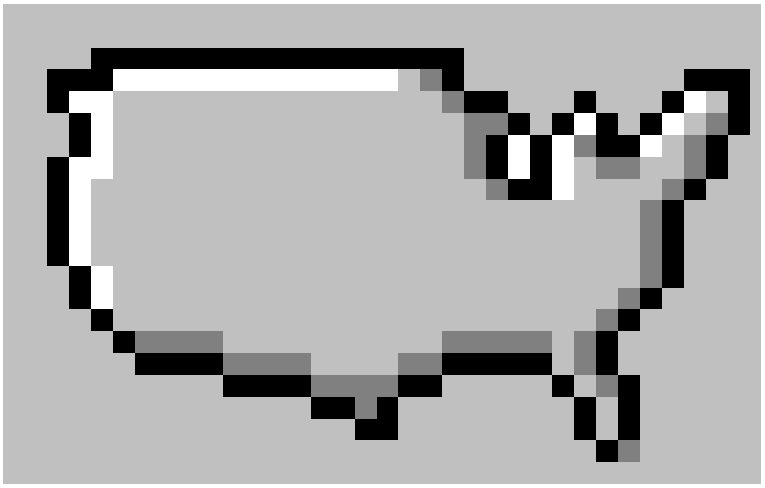
United States



UNEMPLOYMENT INSURANCE DATA VALIDATION HANDBOOK

Contribution Reports

CONTENTS

CONTENTS

EXHIBITS

TAX DATA VALIDATION
GLOSSARY OF TERMS

Data Element Validation (DEV). A validation procedure that involves sorting and reviewing files of reconstructed employer contributions transactions. DEV ensures that individual data elements have been accurately classified for all transactions and match the Federal definitions (Module 2).

File Integrity Validation (FIV). A validation procedure that compares characteristics of reconstructed transactions to supporting documentation (usually query screens on the UI database). FIV ensures that the data in the reconstruction file accurately represent the correct data in the database (Module 2).

Independent Count and Summary Spreadsheet. A validation procedure used when the State has produced the validation file from the same extract files used to produce the ETA 581 Report. Results from independent count validation are entered onto a spreadsheet for comparison with the counts on the Federal reports (Appendices C and D).

Report Validation (RV). A validation methodology that reconstructs (lists) all transactions in mutually exclusive groups (subpopulations), counts them, and compares the results to the counts reported on the Federal Employment Training Administration (ETA) 581 report (Module 1).

Modules. The component processes for conducting data validation. Modules also serve, in practice, as the chapters of this handbook.

Population. A set of transactions of a single type. There are five populations: active employers, report filing, status determinations, accounts receivable and field audits. **Subpopulations** are finer subdivisions of transactions within each population.

RV Files. Files of reconstructed employer contributions transactions created by the State data processing staff using the reconstruction file specifications. Five files are created, one for each population of transactions (Appendix A).

RV Spreadsheets. Spreadsheets that help the State compare subpopulation counts with Federal report item counts for each of the five transaction populations. A separate spreadsheet is available to enter the results of the wage item validation in Module 5 (Appendix B).

Transaction Validation (TV). A validation process that checks the accuracy of each characteristic or data element against all available information in the database, and, if necessary, against information in hard copy documentation. TV has two components — DEV and FIV (Module 2).

Validation Worksheets. Worksheets that guide the validator through FIV and the DEV for each of the five transaction populations (Appendix E).

INTRODUCTION

A. OVERVIEW OF DATA VALIDATION: PRINCIPLES AND APPROACH

Principles

States regularly report to the U.S. Department of Labor (DOL) under the Unemployment Insurance Required Reports (UIRR) system. In particular, States submit a quarterly report on their activities in the collection of Unemployment Insurance (UI) Employer Contributions (taxes). This is the Employment and Training Administration (ETA) 581 report entitled "Contribution Operations."

Data from the ETA 581 report are used for three critical purposes: (1) allocation of UI administrative funding based on State workload, (2) performance measurement to ensure the quality of State Unemployment Insurance program operations, and (3) calculation of State and national economic statistics. Exhibit I.1 summarizes the types and use of the data. Exhibit I.2 displays the ETA 581 report.

EXHIBIT I.1

GENERAL TYPES OF ETA 581 DATA TO BE VALIDATED

Data Type	Funding/ Workload	Performance/Tax Performance System (TPS) Computed Measures	Economic Statistics
Active Employers	✓	✓	✓
Report Filing		✓	
Status Determinations		✓	✓
Accounts Receivable		✓	✓
Field Audits		✓	
Wage Items	✓		✓

Because workload and performance data have these critical uses, it is essential that States report their activities accurately and uniformly. Data validation is intended to assure accurate reporting of employer contribution activities. Two principles underlie a comprehensive data validation process:

EXHIBIT I.2

FORM ETA-581

Contribution Operations

U.S. Department of Labor
Employment and Training Administration

A. Report for quarter ended	B. Region Code	C. State Code	D. State	E. Wage Reporting Basis ("X" one) ☐ Request Reporting ☐ Wage Record	OMB Approval No. 1205-0178 Expires: 8/31/2002

	Line	End of Quarter Employers			Delinquency Cutoff Date	Total Number of Wage Items Received
		1. Contributory	2. Reimbursing	3. Total	4.	5.
Employer Count	101					

		Filing Reports For Preceding Quarters						12. No. of Outstanding Quarters Prior to Report Quarter	13. Total Estimated Contributions Due
		Contributory Employers			Reimbursing Employers				
Employer Reports		6. Filing Timely	7. Secured	8. Resolved	9. Filing Timely	10. Secured	11. Resolved		
	201								

		Status Determinations Made During Report Quarter							
		Newly Established Accounts			Successor Accounts			20. Inactivations/Terminations	
Status Determinations		14. Number	15. Time Lapse of 90 Days or less	16. Time Lapse of 180 Days or Less	17. Number	18. Time Lapse of 90 Days or less	19. Time Lapse of 180 Days or Less		
	301								

Contributory Employers:		21. Total Receivables at Beginning of Period	22. Amount Determined Receivable During Report Period	23. Receivables Liquidated During Report Period	24. Receivables Declared Uncollectible During Report Period	25. Receivables Removed at End of Report Period	26. Total Receivables at End of Report Period	27. No. of Employers Owing Receivables
Receivables	401							
Age of Receivables	402	28. 6 Months or Less	29. 9 Months	30. 12 Months	31. 15 Months		32. Over 15 Months	
Reimbursing Employers:		33. Total Receivables at Beginning of Period	34. Amount Determined Receivable During Report Period	35. Receivables Liquidated During Report Period	36. Receivables Declared Uncollectible During Report Period	37. Receivables Removed at End of Report Period	38. Total Receivables at End of Report Period	39. No. of Employers Owing Receivables
Receivables	403							
Age of Receivables	404	40. 6 Months or Less	41. 9 Months	42. 12 Months	43. 15 Months		44. Over 15 Months	

		Number of Audits			48. Calendar Quarters Audited	Total Wages Audited		51. Hours Spent in Auditing	52. No. of Employees Misclassified as Indep. Contractors
Audit Activity		45. Large Employer Audits	46. Change Audits	47. Total Audits		49. Pre-Audit	50. Post Audit		
	501								

		Amount Underreported			Amount Overreported		
		53. Total Wages	54. Taxable Wages	55. Contributions	56. Total Wages	57. Taxable Wages	58. Contributions
	502						

F. Signature Title Date

Persons are not required to respond to this collection of information unless it displays a currently valid OMB control number. Respondents obligation to reply to these reporting requirements are required to obtain or retain benefits (SSA302(a)). Public reporting burden for this collection of information is estimated to average 8 1/2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the U.S. Department of Labor, Office of Unemployment Insurance, Room S-4516, Washington, DC 20210 (Paperwork Reduction Project 1205-0178).

ETA-581 (Rev. Mar. 1999)

- If data are collected, they should be thoroughly validated to ensure that they are valid and usable.

- If data are collected, they should be thoroughly validated to ensure that they are valid and usable.To be practical for national implementation, the validation approach must be efficient and cost effective.

These two principles have been used to develop this system for validating data that States report to the U.S. Department of Labor about their UI contribution operations. This handbook explains in detail how to do data validation.

Approach

The basic approach used in data validation is to ***reconstruct*** the numbers that should have been reported on the ETA 581 form. Because State unemployment insurance records are highly automated, States can develop computer programs that go through State databases and extract all transactions that should be on the report. The use of automation reduces the burden on validators. Once the transactions are extracted, they are subjected to a series of "logic rules." These rules validate the accuracy of the reconstructed data, assuring that States have used the most definitive source of information and have adhered to Federal definitions.

Modules 1, 2, and 3 of this handbook explain how to create and test reconstruction files — files that list all transactions or employers that should have been counted on the report. The instructions are State-specific in that they present many details of the validation process using the terminology and data elements of each individual state.

Once the State has developed an accurate reconstruction of the numbers that should have been reported on the ETA 581, the handbook provides tools to compare the reconstructed counts to the values that were actually reported.

Modules 4 and 5 of this handbook provide instructions for two validation tasks that do not use an automated reconstruction approach. Module 4 tests the procedures for selecting some samples used in the Tax Performance System. Module 5 describes procedures for validating counts of wage items processed.

B. DATA ERRORS IDENTIFIED THROUGH VALIDATION

Validation is intended to ensure that data on the ETA 581 report are error free. Thus the design of the data validation system should be grounded in an understanding of likely sources of reporting error — both systematic errors and random errors.

Systematic errors involve faulty design or execution of reporting procedures or the automated programs that generate reported counts. Random errors are mistakes in judgment or data entry that corrupt the information entered in data systems or recorded on reports. The validation design addresses both types of error.

- **Systematic errors** can be serious because they are imbedded in automated reporting programs and standard state reporting procedures. Thus, when they occur at all, they occur repeatedly. On a more positive note, the systematic nature of these errors means that they do not need to be assessed very often, and, once corrected, are unlikely to reoccur.
- Systematic errors can produce three types of misreporting: (1) too many transactions reported (overcounts), (2) too few transactions reported (undercounts), and (3) misclassification of transactions. The primary purpose of the data validation process is to identify the occasions when systematic errors produce incorrect reports.
- **Random errors** are more variable. They include problems such as: (1) input errors, (2) judgment errors, for example, misunderstanding or misapplying Federal definitions, and (3) improper State definitions or procedures. In general, random errors occur intermittently. For example, a few data entry errors may occur even when most information is entered correctly. Correcting one error does not ensure that similar errors will not occur in the future.

Many of the more common judgment and definitional errors can be detected through existing Tax Performance System (TPS) reviews. TPS acceptance samples for status determinations and field audits evaluate the accuracy of transaction classification and posting. TPS system reviews look for strong supervision, up-to-date documentation, and other controls that limit human error. Data validation does not attempt to repeat TPS reviews. Rather, it supplements TPS with a review of systematic errors while remaining alert to additional sources of random error.

Consistent and accurate reporting requires both good data and accurate systems for reporting the data. Data validation and TPS together test whether data are entered accurately and whether they are counted correctly. Because TPS tests for many random errors, data validation concentrates primarily on systematic errors.

C. DATA SOURCES FOR FEDERAL REPORTING AND VALIDATION

States use different methods to prepare the ETA 581 report. Some States produce the Federal reports directly from the employer contribution database. Computer programs

scan the entire database to select, classify, and count transactions. Other States produce a database extract or statistical file as transactions are processed, essentially keeping a running count of items to be tabulated for the report. Still other States use a combination of these methods. The basic approach to Data Validation is the same no matter how the report is developed — States reconstruct the reported transactions using standard national criteria.

The validation methodology is flexible in accommodating the different approaches used by States. However, validation is most effective when validation data are produced directly from the employer contributions database. For cost reasons and to minimize changes in data over time, some States prefer to use daily, weekly, or monthly statistical extract files instead. When extract files are used, other types of system error may occur. Reportable transactions may be improperly excluded from the employer master file. Furthermore, the statistical file may contain corrupt data. The statistical file is not used as part of the daily tax system and, therefore, errors may not be detected and corrected through routine agency business.

The only way to test for these problems is to independently reconstruct or query the employer master file. States that produce validation data from the same extract files used to produce the ETA 581, rather than directly from the database, must ensure that the extract files contain all the appropriate employers or transactions. The way to do this is to recreate the logic used to produce the reports. This handbook includes a validation tool, "independent count validation," specifically for this purpose.

The specific type of independent count (simple query, multiple queries, cross tabulation) must be determined by State programming staff.[1]

Exhibit I.3 outlines variations in the validation methodology, based on typical State approaches to ETA 581 reporting and data validation reconstruction. To identify the specific validation methodology to be implemented, the State validator or regional representative should identify the State's ETA 581 report source and validation reconstruction source for each population to be validated.

[1] There is no way to accurately reconstruct the report count when the statistical file contains transactions that are no longer present on the database (e.g., when it still includes status determinations that were deleted from the main database after a corrected status determination was prepared for the same employer). The State may only be able to do a rough independent count, e.g., by comparing the count of new status determinations to the difference between the prior and the report quarter's count of active employers.

EXHIBIT I.3

VARIATIONS IN VALIDATION METHODOLOGIES BASED ON STATE APPROACHES TO REPORTING AND RECONSTRUCTION

Scenario	Transactions Overwritten on Database	ETA 581			Data Validation			Independent Count Required	Source Documentation Review Required	Comments
		Program Type	Source	Timing	Program Type	Source	Timing			
1	No	Count	Database	Snapshot	DRE	Database	Snapshot	No	No	Best scenario because comparing snapshots eliminates timing discrepancies
2	No	Count	Statistical file	Daily	DRE	Database	Snapshot	No	No	Database is only reconstruction source. There could be changes in transaction characteristics (but will find all transactions).
3	No	DRE	Database	Snapshot	DRE	Database	Snapshot	Yes	No	Reporting and validation are the same program. Independent count may mirror that program.
4	No	DRE	Statistical file	Daily	DRE	Statistical file	Daily	Yes	Yes	Since transactions are not overwritten, States should be able to do Scenario 2 instead.
5	Yes	DRE	Statistical file	Daily	DRE	Statistical file	Daily	NA	NA	No alternative validation source. Cannot reconstruct from the database. Not thorough validation.
6	Yes	Count	Statistical file	Daily	Must create a daily extract	NA	NA	NA	NA	Cannot reconstruct from database. Must change reporting process to Scenario 5.

NOTE: Snapshot is of a reporting period.
DRE = Detail Record Extract
NA = Not Available

D. VALIDATION APPROACH

The comprehensive data validation program described in this handbook ensures the accuracy of the UIRR data. It validates most items on the ETA 581 report. It does so using a process that is highly automated and complements existing quality control programs such as TPS. Thus it minimize validator time and ongoing State burden.

The validation process involves a reconstruction of report counts. It provides an audit trail to support the counts and classifications of transactions. Through this audit trail, the State proves that its UIRR data have been correctly counted and reported. For example, if a State reports 5,000 reimbursable employers at the end of the quarter, then the State must create a file listing all 5,000 employers. The file also lists relevant characteristics, such as the Employer Account Number (EAN), employer type, liability date, number of liable quarters, and sum of the wages in those quarters. Analysis of these characteristics can assure validators that all 5,000 employers belong on the list. If such a file does, indeed, contain 5,000 correctly classified employers then the reported number is proved and the report is considered valid.

To assure that the reconstruction of report counts has been done correctly, the approach also includes three tests of the reconstruction process:

1. Module 1.2, Duplicate Detection, checks that each transaction is included in the reconstruction only once.

2. Module 2.1, File Integrity Validation, ensures that the correct data has been extracted from the State's database.

3. Module 2.2, Data Element Validation, ensures that each transaction in the reconstruction file is classified correctly.

These three checks build validators' confidence in the reconstruction count.

This handbook provides detailed validation instructions for each State, ensuring that State and Federal Regional Office staff understand all relevant aspects of the State's employer contributions reporting system. In specifying how to reconstruct reported transactions, the methodology explains the criteria that States should use in their Federal employer contributions reporting. Thus in addition to guiding the States through the validation process, this handbook provides technical guidance on Federal ETA 581 reporting requirements

States themselves perform the validation, which is followed by a Federal Regional Office audit of the results. To facilitate the Regional Office audit, the State prepares

and maintains a validation package. This package enables the regional auditor to easily follow the validator's work, without requiring the State to print out entire reconstruction files. The components of the package are discussed in more detail in Modules 1 and 2.

E. UNITS OF ANALYSIS

There are 50 ETA 581 report items to validate.[2] Each item has its own set of definitions, rules, and validation requirements. To minimize the burden of reconstructing item counts, the validation process is streamlined by breaking it down into manageable components. The data to be validated are grouped into mutually exclusive ***populations*** and ***subpopulations***. The validation process itself is organized into a series of ***modules*** that are the logical steps in the process. This structure lets the validation team focus on one type of data at a time, and validate each type of data one step at a time.

A single employer account transaction may be counted in several different ETA 581 report items. For example, a timely filed contribution report is counted in two items for the current report quarter (timely reports and reports secured) and in one item in the following report quarter (reports resolved).

Validation reconstructs and analyzes each transaction only once, even if it is counted in multiple cells on the report. Employers or transactions are classified into mutually exclusive groups — specifically, five types of employers or transactions (populations), which are composed of 46 mutually exclusive groups (subpopulations). Each subpopulation represents a unique set of data elements or characteristics. All validation counts are built from these subpopulations. The five populations of reconstructed employer contributions transactions are:

1. Active employers
2. Report filing
3. Status determinations
4. Accounts receivable
5. Field audits

[2] Wage items processed (item 5 on the ETA 581) are validated but through a less comprehensive process. They are not included as a reconstruction population.

Exhibit I.4 lists the ETA 581 populations and subpopulations that are reconstructed and the number of report items being validated for each.

EXHIBIT I.4

ETA 581 REPORT, BY TRANSACTION POPULATION

Transaction Population	ETA 581 Line Numbers	Dimensions Used to Distinguish Subpopulations	Number of Report Items	Number of Subpopulations
1. Active Employers	101	Employer status • contributory • reimbursing	3	2
2. Report Filing	201	Timing of report receipt and resolution • timely • secured within the quarter • resolved within two quarters	6	16
3. Status Determinations	301	Type of status determination • new • successor • inactive • terminated Time lapse of the determination	7	8
4. Accounts Receivable	401 402 403 404	Type of receivable processing • amounts established • liquidated • declared uncollectible • removed from the report • outstanding debt.	22	16
5. Field Audits	501 502	Employer size • small • large Audit result • change • no change	11	4
Wage Items Processed	101		1	N/A

F. HANDBOOK OVERVIEW

To ensure that reported data are accurate and meet Federal reporting definitions, there are five validation processes or "modules". These modules provide all the tools to be used in validating the quantity and quality of Federally reported data. The modules are outlined below.

Modules

- **Module 1 — Report Validation (RV)**

  Module 1 validates that the programs that create the Federal reports are functioning correctly. The module provides instructions for creating the extract files that are audit trails for information in the ETA 581 report. Detailed specifications for these extract files can be found in Appendix A. The validator systematically examines the reconstructed files once they are produced. The following tools are used to accomplish this validation. They are introduced briefly here and discussed more fully in later sections.

  (1.1) Item Count. The validator compares the count in each Federal report item with the count from its corresponding subpopulations in the reconstruction files. An RV spreadsheet is provided to perform all necessary calculations and determine whether any differences between the counts are within an acceptable level of error.

  (1.2) Duplicate Detection. Samples of reconstructed transactions are sorted by employer account number (EAN). The validator reviews these files to detect any invalid duplicate transactions. The results are then entered onto the RV spreadsheet and the final reconstruction counts are adjusted for duplicates as necessary.

- **Module 2 — Transaction Validation (TV)**

  Module 2 validates individual transactions to determine the accuracy of both the State reconstruction files and the data elements used to classify the transactions. Validators use the reconstruction files generated for Module 1 along with the sort files created for Module 2. Each column in the file refers the validator to a specific "step" or substep in the State-specific validation instructions (Module 3). The instructions refer the validator to the appropriate screen, field, and code used to validate the data item in that State. The validator tests this information for selected transactions to ensure that

the reconstruction files have been built correctly. The following tools are used to accomplish this transaction validation:

(2.1) File Integrity Validation. For a minimum sample of two (2) transactions, the validator compares characteristics of the transaction to all available supporting documentation in the State's records. This ensures that the data in the reconstruction file accurately reflect the correct employer records in the database. Results are entered on the validation worksheet.

(2.2) Data Element Validation. For specified data elements, the validator reviews sorted reconstruction files to ensure that all transactions have been accurately classified based on the individual data elements and the Federal definitions. Results are entered on the RV spreadsheet and the validation worksheet.

- **Module 3 — State-Specific TV Instructions**

  This is the State-specific set of instructions that the validator uses in TV.

- **Module 4 — TPS Validation**

  This module describes a process to validate TPS acceptance samples.

- **Module 5 — Wage Item Validation**

  This module explains how wage items are to be validated.

Appendices

- **Appendix A — RV Specifications** contains detailed requirements for producing the reconstruction files for Module 1 and the sorts used in Module 2.

- **Appendix B — RV Spreadsheets** provides forms for reporting the results for each validated population.

- **Appendix C — Independent Count** provides a mechanism for the validator to determine whether any transactions have been excluded from any ETA 581 report item. This mechanism is applicable to States that create the ETA 581 from the same extract files used to generate the reconstructed files. It is not required for States that use separate programs to generate Federal reports and to reconstruct the reported transactions.

- **Appendix D — Independent Count Summary Spreadsheet** is used to record the results of an Independent Counts. It is used only by states required to conduct an Independent Count for validation.

- **Appendix E — Validation Worksheets** guides the validator through the TV process. These worksheets record any errors identified through TV.

- **Appendix F — Alternative Methodology For Duplicate Detection** describes a more automated test for duplicates that States have the option of using.

- **Appendix G — Data Validation Tasks** provides a summary of recommended staff roles and activities for completing data validation.

G. WALKTHROUGH OF THE DATA VALIDATION METHODOLOGY

Exhibit I.5 illustrates the data validation process detailed in the handbook modules and appendixes, using ETA 581 active employers as an example.[3]

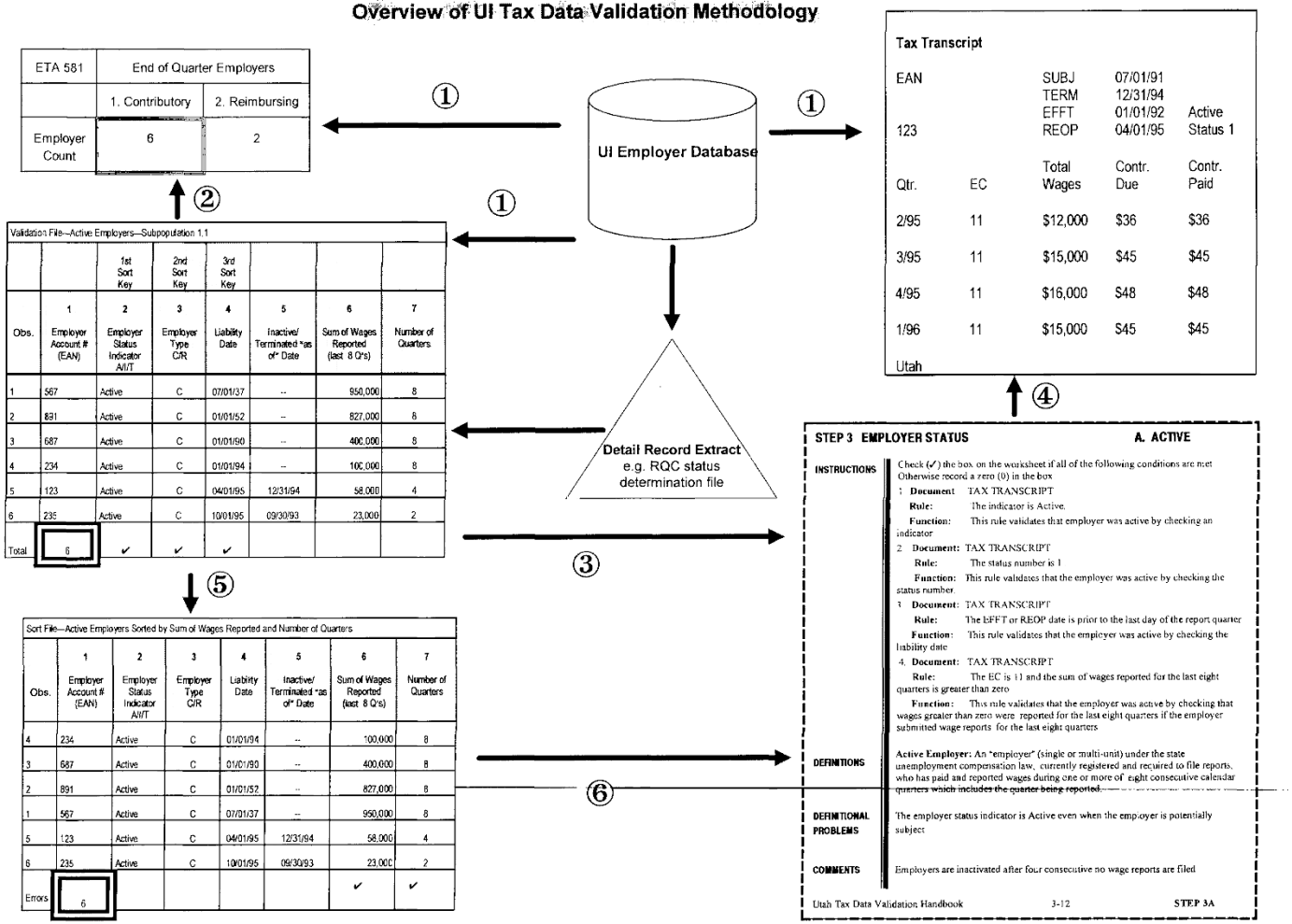
① The "***UI Employer Database***" represents the State's master file, or perhaps several databases storing data on different parts of the tax operation. As States manage employer contributions, they enter data in the database. There may also be times when the system automatically places data in employer accounts, for example, when an automated flag is placed in an employer file to indicate that a report is delinquent.

As Exhibit I.5 shows graphically, the State may view data from the database in several different ways. A State may query the database, for example, by referring to a query screen such as the "Tax Transcript" at the top right of the figure. The State may also produce more formal reports such as the ETA 581, a portion of which is shown at the top left.

② To validate the official report, the ***state data validation staff*** develops a detailed record extract, or ***reconstruction file*** — a list of all transactions on the state database that should be reported for a single item on the ETA 581. The file is displayed for review by validators (middle figure in left column).

[3] The validation file, sort file, and State-specific handbook have been modified slightly in Exhibit I.5 for presentation purposes. Utah's Tax Transcript screen and handbook are shown.

Exhibit I.5
Overview of UI Tax Data Validation Methodology



The ***validator*** compares the reconstructed count (here 6) to the reported count (also 6). Transactions are also checked for duplicates and other errors.

The State should concurrently generate the ETA 581 and the validation file (the reconstructed "audit trail") from the employer database. At the same time the State should produce supporting documentation from the employer database (for example, query screens) for File Integrity Validation (FIV) (Module 2).[4] Alternatively, the State may generate the ETA 581 and/or the validation file from a detail record extract statistical file (such as a TPS universe).

The ***validator*** assembles a package of materials — printouts of the beginning and end of each reconstruction and sort file, population spreadsheets, worksheets, and screens — to be used during validation and for review by an auditor from the Regional Office of the U.S. Department of Labor.

③ The ***validator***, guided by "Step" numbers in each column heading on the reconstruction specification, refers to instructions in the State-specific portion of the validation handbook to test the integrity of the reconstructed data. The bottom right portion of Exhibit I.5 shows a page of Module 3. The "Steps" in Module 3 provide State-specific instructions for checking that the reconstruction files have been built correctly.

④ The handbook refers to State source documentation (usually query screens) that the validator compares to the reconstruction file to complete FIV.

⑤ To perform Data Element Validation (DEV), ***State ADP staff*** sort, by individual data elements, all transactions in specified subpopulations from the validation reconstruction file. The sorted files help the validator spot transactions that should not have been included in the reconstruction files because they do not meet the specifications for the subpopulation.

⑥ The validator, again guided by step numbers in each column heading, refers to specific steps in the State-specific validation handbook to validate that the

[4] Given the highly automated nature of tax data validation, database screens are generally the only supporting documentation needed. Therefore, this handbook refers to screens, rather than to supporting documentation, throughout. To prevent inconsistencies due to timing, screens to validate the accuracy of transfer of data from the database should be printed as the reconstruction file is created. If this is not possible, the validator may select non-overwritten transactions for the FIV.

individual data elements have been accurately classified and match the Federal definitions.

⑦ If necessary, after reconstruction files have been tested and corrected, the validator makes a final comparison between the reconstruction counts and the report counts.

H. OVERVIEW: PREPARATION FOR DATA VALIDATION — TASK 1

From time to time this handbook provides ***recommendations*** for managing the validation process. These recommendations are described in "Task" exhibits. For each task the handbook provides a listing of activities to be completed and the staff who are likely to take the lead on each activity. Staff roles and responsibilities for preparing for Data Validation are summarized in Exhibit I.6 below. Staff roles will be divided among:

- Automated data processing (ADP) staff, who have the primary responsibility for extracting data from the database to create the reconstruction files. They also sort and format those files so they are useful to validators.

- Validators, the end users who test the reconstructed data and then assess the validity of the information the State has reported on the ETA 581 report. Validators should work closely with ADP staff to determine the information that belongs in the reconstruction files.

- Managers, who are responsible for assuring that (1) the data validation process stays on track and (2) the data validation team has the resources it needs to meet the requirements of this handbook and the schedule set by the State.

EXHIBIT I.6 TASK 1: PREPARING FOR DATA VALIDATION	
Activity	**Roles**
Assemble data validation team.	Managers
Review handbook.	Validators, ADP
Attend training. Share training with staff who did not attend.	Validators, ADP
Review and update State-specific information in Module 3 of the handbook. Send U.S. DOL and its contractor a copy of the module with any needed changes clearly marked.	ADP, with help from validators
Develop a data validation plan with: • Schedule for completing data processing and validation review ***for each population.*** • Staff assignments for each step in the data validation process.	Managers, validators, ADP

Module 1

REPORT VALIDATION

The remainder of this handbook guides users through the data validation process. Modules 1, 2, and 3 describe the major steps required to conduct data validation. The appendices provide the forms and specifications needed as the State proceeds through these steps.

Module 1 presents the process for:

- Producing reconstruction files,
- Counting the transactions or employers in each file, and
- Comparing the reconstruction counts to the counts that appeared on the ETA 581 report.

Module 1.1, Item Count, counts the transactions that should be reported and compares them to the transaction counts the State has actually reported. The comparison between the reconstruction count and the reported count determines whether the State reporting process is judged to be valid.

Module 1.2, Duplicate Detection, identifies duplicate records in the State's UI tax database. If duplicates are found, adjustments are made so that the reconstruction is based on an unduplicated count of transactions.

The duplicate detection submodule also begins a three-step process of testing the reconstruction file.

- ***Module 1.2, Duplicate Detection***, screens for duplicates to ensure that the reconstruction counts each transaction only once.
- ***Module 2.1, File Integrity Validation***, checks that the correct data elements have been used when building the reconstruction files.
- ***Module 2.2, Data Element Validation***, examines whether all transactions assigned to a subpopulation truly meet the parameters for inclusion in that subpopulation.

Once these quality checks are complete, the validator can be certain that the reconstructed count is the correct value to compare to the reported count.

MODULE 1.1 — ITEM COUNT

A. PURPOSE

This procedure validates the accuracy of the counts reported for items on the ETA 581. Five Report Validation (RV) files are produced. These files reconstruct the counts for the five different types of transactions being validated. The ***validation counts*** are used to determine the accuracy of ETA 581 ***report counts***. The five RV populations are listed in Exhibit 1.1.

EXHIBIT 1.1

SUMMARY OF REPORT VALIDATION POPULATION FILES

File Specification	Population	ETA 581 Line Number
1	Active employers	101
2	Report filing	201
3	Status determinations	301
4	Accounts receivable	401, 402, 403, 404
5	Field audits	501, 502

B. METHODOLOGY

Appendix A, Part I, guides the automated data processing (ADP) staff through the programming logic used to create the validation files for each population. This appendix contains detailed specifications for producing five RV files. It includes the criteria for organizing populations into their component subpopulations. It also shows sort specifications giving the order for listing transactions in each RV file. Sort specifications are also provided for use in Module 2.2 (Data Element Validation).

1. RV File Specifications

Exhibit 1.2 is a copy of the first page of the reconstruction file specifications for population 1, Active Employers.[1] It may be helpful to walk through the key features of the specifications, by the numbers.

[1] The entire specification can be found in Appendix A, page A.3.

① Appendix A, Part I, includes a table for each of five transaction populations.

Example: This table is for population 1, active employers.

② Each transaction population is further divided into subpopulations to match the types of transactions that are reported on ETA 581. For example, population 1 is subdivided into two subpopulations by type of employer (contributory and reimbursing). The first column in the table lists the subpopulation numbers.

Each row in the table provides the specifications for an entire reconstruction file for the subpopulation. This will be a portion of the reconstruction file for the whole population.

Example: This row tells ADP staff to create a list of all contributory employers that were liable at the end of the report quarter (subpopulation 1.1).

③ A verbal description of each subpopulation follows each table in Appendix A.

④ The second column in the specification indicates the ETA 581 item(s) to which this subpopulation count is compared. In population 1, there is a one-to-one match between subpopulation counts and report items. However, in other populations one subpopulation may be used in the validation of two or more report items.

Example. In population 2, found on page A.5, the validation count for Total Reports Secured is calculated by adding together validation counts for two subpopulations — subpopulation 2.1 (reports received timely) and, subpopulation 2.2 (non-timely reports secured by the end of the report quarter). This sum, is compared to the value found on ETA 581, Item 7, secured reports. (The validation count for subpopulation 2.1, by itself, is also the comparison value for ETA 581 item 6, reports received timely.)

① **Table 1 Draft Reconstruction File Specifications**
Active Employers

Report Quarter:
Validator:
Date:

These subpopulations constitute the unique subgroups of all active employers on the last day of the Report Quarter (RQ) covered by the ETA 581. Reconstruction should be done at the end of the RQ being validated (when the ETA 581 report program is run).

⑦

			1st Sort Key	2nd Sort Key	3rd Sort Key	4th Sort Key				
⑤ Sub-population	④ Reported in 581 Item #'s	1 *(Step 1A)* Employer Account # (EAN)	2 *(Step 3A)* Employer Status Indicator A/I/T	3 *(Step 2A)* *(Step 2B)* Employer Type C/R	4 *(Step 4B)* Liability Date (Initial)	5 *(Step 4A)* Liability Date (Reopen)	6 *(Step 5)* Inactive/ Terminated "as of" Date	7 *(Step 15)* Activation Processing Date	8 *(Step 7B)* Number of Liable Quarters	9 *(Step 7A)* Sum of Wages (Last 8 Q's)
② 1.1	1	⑥➜	A	C	<=RQ	<=RQ	>RQ, or <liability date (reopen), or none			(If col. 8≥8) >$0
1.2	2		A	R	<=RQ	<=RQ	>RQ, or <liability date (reopen), or none			(If col. 8≥8) >$0

Notes:

1) The 3rd and 4th sort keys should be treated as a single sort. Sort by the most recent date.
30. Either column 4 or 5 must be a date earlier than the end of the report quarter.

③ Subpopulation descriptions:

1.1 Active contributory employers liable by the end of the report quarter.
1.2 Active reimbursable employers liable by the end of the report quarter.

⑤ The column headings in the specification tables serve two purposes:

a. They make it easier for validators and regional auditors to move between the printouts and the instructions in this handbook; both have the same headings. The column headings, should be printed on the reconstruction file exactly as they are in the specifications — including steps, column numbers, and column labels.

b. They list all the data elements that should be displayed for each transaction (or employer) in the population.

Example: Each line in the reconstruction file for population 1, active employers, should show the subpopulation number, the employer account number, the status of the employer (active/inactive/terminated), the employer type (contributory/reimbursing), and so on.

State ADP staff should program the RV files so that the generic terms used in Appendix A are displayed, as well as the actual codes and values from the State files. For example, if the State's code for contributory employers is "1," the programmer should have the RV file list "Contributory-1" or "C-1" under employer type for all contributory employer subpopulations. When a State has multiple codes for a single Federal reporting criterion, the file should list the specific codes that apply to each transaction. For example, if the State maintains two codes for contributory employers, "1" and "2," the programmer should list "C-1" or "C-2," as appropriate.[2]

Displaying the generic reporting term or an understandable abbreviation of the term for each record, in addition to the actual data value, facilitates validation and permits greater flexibility in staffing the effort. No translation is required for new staff to read the data output and locate the corresponding validation instructions.

In addition to the columns shown in the specification, all RV files should include an ***observation number*** to the left of each employer or transaction. The observation number should begin with 1 for the first transaction on the list and continue consecutively to the end of each ***subpopulation***. Having the observation number makes it easy for validators to know how many transactions are in each file. The observation number also helps validators when they need to check the data in a specific transaction. They can do their

[2] To the extent that a State has multiple codes for a value, the validator may need to review additional transactions to check transaction parameters in DEV in Module 2.2.

research and then, using the observation number, quickly return to the same transaction in the printout.

⑥ In the table, the non-blank columns provide the specifications for which transactions or employers should be included in the subpopulation.

Example: To be included in subpopulation 1.1 an employer must:

(a) be Active at the end of the quarter,
(b) be a Contributory employer,
(c, d) have an initial liability or reactivation (reopen) date prior to the end of the report quarter (RQ),
(e) not have a termination date unless it either preceded the most recent reopen date or did not occur until the end of the report quarter ,
(f, g) not have filed reports for eight or more consecutive liable quarters, and reported zero wages for the last eight of them.

⑦ The reconstruction file should be sorted by the **Sort Keys** in the order they are numbered.

Example: The reconstruction file for population 1 is sorted first by the employer status indicator, then by employer type, then by liability or reopen date (whichever applies).

2. Report Quarter Terminology

The specifications in this handbook use a shorthand terminology to refer to report quarters. Exhibit 1.3 is a time line illustrating how terms and symbols are used.

- The **Report Quarter (RQ)** is the time period shown on ETA 581 in the block labeled "A. Report for quarter ended." This means that the ETA 581 report is showing transactions that occurred during the quarter or the status of transactions as of the end of the quarter. For example, the ETA 581 report includes items such as the number of active employers at the end of the report quarter and the number of timely employer reports received during the RQ. The report quarter ends at point A. Point A is also the time when the State runs programs to download data for both the ETA 581 counts and the data validation reconstruction files. The ETA report which relates to the report quarter is due at the hashmark labeled "ETA 581 Due" in Exhibit 1.3

- Contribution and wage reports received from employers during the Report Quarter reflect employer activity that occurred during the quarter before the report quarter (**RQ-1**). Because this prior quarter is the subject of employer

EXHIBIT 1.3

QUARTERLY TIMELINE

2001						2002					
3			4			1			2		
Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun

RQ-1 (ERQ)
- Reports delinquent in this quarter can be received by RQ+1

RQ: Report Quarter

States:
- register employers
- determine their liability status
- receive reports
- pursue delinquent reports
- Establish, liquidate, and write off receivables
- conduct field audits

RQ+ 1

States:
- receive reports on activity in RQ
- continue resolving RQ-1 reports

Employer Reports Due (for RQ-2)

Employer Reports Due (for RQ-1)

ETA 581 Due (for RQ)

ETA 581 Due (for RQ+1)

A

B

A **Data Processing Staff: Product Report Counts**
Complete Validation Extract
Prepare Screen Prints

B **Validation files and sorts prepared and formatted for review**
Validation reviews begin (time is approximate)

reports received during the RQ, RQ-1 is sometimes referred to as **ERQ** or **Employer Report Quarter.** When specifications need to refer to earlier quarters, they will extend the basic convention. The quarter prior to RQ-1 is RQ-2, the quarter prior to that is RQ-3, and so on.

- The specifications refer to the quarter after the report quarter using the term **RQ+1**. This term is used most often for population 3, report filing, where States have through RQ+1 to resolve reports due in RQ.

3. Issues to Consider When Producing RV Files

Timing. The ETA-581 report is a snapshot of performance at one point in time. Like the well-advertised "Kodak moment," if you miss the moment, the picture may be gone. Depending on the State's system, employer records may always be changing as employers are terminated or added to the rolls, payments and adjustments modify account balances, long-delinquent reports finally show up, and so forth. When this happens State data systems may overwrite earlier records, making them inaccessible.

States should produce the RV reconstruction files at the same time they produce the ETA 581 report. This will eliminate the chance that validation counts will differ from the report simply because transactions were added to or removed from the employer account in the interval between running the ETA 581 report and the data validation reconstruction. If there is a slight difference in the timing of the two runs, the data "as of" the time when the ETA 581 report was run can sometimes be reconstructed if the State has a complete audit trail. Theoretically the validator could use the audit trail to verify that a transaction was correct at the time of reporting. However, this would be cumbersome. It is better to download data simultaneously for reconstruction files and the ETA 581 report.

Capture files. There is another, more troublesome problem related to the production of the RV files. The problem is that some transactions may overwritten or changed in some State databases. Thus, if the original record has not been captured and saved, it will disappear and will not be available to validators. In data validation, overwritten transactions are a problem primarily in two populations:

- ***Status determinations (population 3)*** where, for example, an employer might be registered as newly liable. After investigation of the case the employer record is modified to indicate that it should have been classified as a successor, taking over an existing business. The record of the first status

determination, the one classifying the employer as newly liable, is erased and replaced.[3]

- ***Accounts receivable (population 4)*** where payments and adjustments can change the values in employer accounts, overwriting the prior account status and making that status information accessible only by referring to internal audit trails.

It may be necessary to create special capture files to ensure that the reconstruction files have access to all transactions. These new files will capture and store all transactions for each employer account. By capturing this information, States can maintain records of status determinations or other overwritten transaction types, even if the transactions are later canceled or adjusted.

If States need to develop special files to capture an audit trail of all transactions, the capture program must run for one full reporting period before validation can be done.

Listing the reconstruction files. Although States are encouraged to produce complete printouts of each reconstruction file, the State validator may choose to review the lists on-line. For example, in large States and for active employers, the reported counts are so high that it is not practical to print the entire population. The largest subpopulations can be viewed on a screen.[4] If on-line review is selected, the State must print at least the first two and last two pages of each the reconstruction file to enable the Federal validation auditor to conduct a review without having on-line access.

[3] The information probably is retained in an employer history file but is more difficult to reconstruct.

[4] Large States have several options in these situations. The best is to create a file containing the validation data but not print it. Validators and auditors can scroll through the file on-line to verify counts and characteristics of transactions. States can also run their Federal reporting programs for a limited period and validate that period's transactions (although this reduces the effectiveness of the duplicate detection procedure). Before validation, States must document how they will validate large populations. ***They must obtain approval from Federal regional representatives before deviating from standard procedures***.

In any case, supporting documentation (screens) should be printed when the programmer creates the RV file so the validator can be sure that the RV files have been constructed properly.[5] (See Module 2 for a detailed explanation.)

OVERVIEW: Module 1.1

At this point in Module 1, you have completed the instructions for producing report validation (RV) files. Before continuing with the discussion of Module 1, it may be helpful to review the steps the validation team will need to complete as the RV files are constructed. That review is presented as Exhibit 1.4.

EXHIBIT 1.4

TASK 2: CREATING RECONSTRUCTION (RV) FILES

Activity	Roles
Review specifications, product requirements, and schedules.	ADP, Validators
Convert handbook specifications into programming specifications.	ADP
Develop "capture" programs if needed.	ADP
Develop file extract programs.	ADP
Develop programs to format extracted files for reviewers.	ADP
Review test output for consistency with handbook requirements.	Validators
Modify programs to correct any problems identified.	ADP
Schedule data validation extract programs to run at the same time ETA 581 programs are executed. Also arrange to print any screens needed for Module 2 at the same time.	Managers, ADP, Validators

[5] If Module 1 validation counts differ substantially from reported counts, a programming error probably has occurred. In this case, research should be conducted to determine if the error is caused by the validation software or the reporting software. Significant errors in the validation software should be corrected before continuing with the validation process. Module 2 describes File Integrity Validation (FIV) and Data Element Validation (DEV), which also confirm that the validation output was programmed correctly.

C. Recording Data Validation Findings

1. Report Validation (RV) Spreadsheets

Five spreadsheets are used to enter validation findings. They are available to States in either paper or electronic forms. The spreadsheets also serve as the reporting vehicle for Data Validation.

Once the correct reconstruction/validation count is identified for each subpopulation, the value is entered on the RV spreadsheet along with the reported count (see Appendix B). The spreadsheets compare subpopulation counts with Federal report item counts and the sort counts for each of the five transaction populations.

On the RV spreadsheets, State staff download or key-enter the applicable subpopulation counts from the RV files and the sort counts from the sort files in Module 2.[6] The spreadsheet performs all necessary calculations and identifies any significant discrepancies. Validated and reported dollar figures are also entered for applicable subpopulations. The spreadsheet groups errors by category, to facilitate analysis of problems, and provides a format for reporting validation errors, or absence thereof, to the U. S. Department of Labor.

2. Instructions for Completing the RV Spreadsheets

a.. In the column labeled ***ETA Reported Count*** enter the reported count for each ETA 581 item.

b. Enter the counts from each subpopulation on the RV file in the ***Validation Count*** column of the RV spreadsheet. Where applicable, enter the dollars from each subpopulation in the ***Validation $*** column of the RV spreadsheet.[7] This can be done by the validator or by a computer program that transfers the subpopulation counts and dollars from the validation file to the spreadsheet.

c. Enter the number of ***transactions reviewed*** and the ***# of duplicates*** detected in the box labeled Duplicate Calculations. The spreadsheet will calculate the

[6] States should enter their final results on the spreadsheet but treat them as ***provisional*** until the they have completed Module 2, verifying the accuracy of the reconstruction. The provisional entry will identify any large differences between validation counts and reported counts, thereby alerting validators to potential errors in the validation software. Once module 2 is completed and the validation data is fully checked, the validator will return to the spreadsheet to enter final values.

[7] The dollar amounts entered on the spreadsheet are the sum of all amounts listed in a column for all subpopulations in the reconstruction file.

percent of duplicates and automatically adjust the validation count for the population as a whole. (See Module 1.2.)[8]

d. When there is not a one-to-one relationship between the validation counts and the ETA 581 counts, the spreadsheets automatically add or subtract validation counts as necessary to make the proper match. For example, subpopulations 2.1 through 2.8 must be combined to match the reported count for resolved reports for contributory employers during a report quarter (ETA 581 item 8).

e. The ***Count Difference*** and the ***Count % Difference*** between the count from the ETA 581 report and the comparable validation count(s) are automatically calculated at the subpopulation and the population levels. If the Count % Difference is greater than plus or minus 2 percent, the Count Pass/Fail column will indicate "Fail."

D. EXAMPLE — RECORDING DATA VALIDATION FINDINGS

Exhibit 1.5 is a sample RV file for population 1. Population 1 has two subpopulations and the RV file shows the transactions (employers) that the programmer included in subpopulations 1.1 and 1.2. Subpopulation 1.1 represents active contributory employers and subpopulation 1.2, represents active reimbursing employers. There are six valid transactions listed in subpopulation 1.1 and two valid transactions in 1.2.

Exhibit 1.6 is a sample RV spreadsheet for population 1. Here the validator has entered the validation counts of 6 for subpopulation 1.1 and 2 for subpopulation 1.2 and the reported counts of 6 for ETA 581 item 1 and 2 for ETA 581 item 2. When the validation counts are compared to the reported counts for items 1 and 2 on the ETA 581, the spreadsheet calculates that there is a count difference of zero percent, and displays, a "Pass."

[8] Checks for duplicates are conducted for the population as a whole, not for subpopulations.

EXHIBIT 1.5

VALIDATION FILE — ACTIVE EMPLOYERS — POPULATION 1

Obs.	1 Employer Account # (EAN)	2 Employer Status Indicator A/I/T	3 Employer Type C/R	4 Liability Date (Reopen)	5 Liability Date (Initial)	6 Inactive/ Terminated "as of" Date	7 Activation Processing Date	8 Number of Liable Quarters	9 Sum of Wages (Past 8 Q's)
1	567	Active	C		07/01/37			8	827,000
2	891	Active	C	3/31/94	01/01/52	9/30/93		8	400,000
3	687	Active	C		01/01/90			2	23,000
4	234	Active	C		01/01/94			4	58,000
5	123	Active	C		04/01/95			8	950,000
6	235	Active	C		10/01/95			8	100,000
Subtotal	6								
1	898	Active	R		04/01/92			7	62,000
2	747	Active	R		10/31/95			8	832,000
Subtotal	2								

EXHIBIT 1.6

REPORT VALIDATION SUMMARY SPREADSHEET — POPULATION 1

	ETA 581 ITEM	ETA REPORTED COUNT	SUBPOPULATION	SUBPOPULATION VALIDATION COUNT	% DUPL. ERRORS	ADJUSTED VALIDATION COUNT	COUNT DIFFERENCE	COUNT % DIFFERENCE	COUNT PASS/ FAIL
Active contributory employers	1	6	1.1	6			0	0.00%	
Active reimbursing employers	2	2	1.2	2			0	0.00%	
All active employers	3	8	TOTAL	**8**	**0.00%**	**8.00**	0	**0.00%**	**PASS**

Duplicate Calculations	
Transactions Reviewed	# of Duplicates
6	**0**

SORT NUMBER	SORT COUNT	SORT COUNT DIFFERENCE
1.S.1		8
1.S.2		8
1.S.3		8
1.S.4		8
1.A		6
1.B		2
1.C		8
1.D		8

ITEM
Active contributory employers
Active reimbursing employers
All active employers

E. FINAL RESULTS

The RV spreadsheet documents, by type of error, discrepancies among the RV files, the Federal ETA 581 report, and the sorts. This allows the validator to identify trends and systematic errors. The validator should further research any "Fail" indicators to determine the source of errors and should document findings in the Comments column on the RV spreadsheet.

See the Results section of Module 2.2 for a discussion of required ***corrective action*** when validation identifies errors.

MODULE 1.2 — DUPLICATE DETECTION

A. PURPOSE

Report validation (RV) files that are produced according to the specifications in this handbook should provide a *single count* of all transactions. Each transaction that is to be reported on the ETA 581 should be included only once in the reconstruction count. However, caution dictates that States test their reconstruction files to make sure that duplicates are not inadvertently included.

Module 1.2 identifies transactions that appear more than once in the State validation count. If any inappropriate duplicates are detected, an adjustment is made to the reconstruction count to eliminate any resulting overcounting. The primary purpose of the module is to detect duplicates in the State database. However, it may also help correct for any duplicates introduced as a result of incorrect programs used when producing the extract files.

In the duplicate detection process a sample of 400 employers is examined to ensure that there are no duplicate employer transactions. Transactions in the sample are sorted by Employer Account Number (EAN) to detect duplicate entries caused by data entry or program errors. Without this sorting, duplicate entries in the RV files would be difficult to identify because they could appear anywhere in the file. The sorts specified in this module for populations 1, 2, 3 and 5 enable the validator to quickly analyze the transactions associated with each EAN because any duplicate transactions will be listed next to each other.[9] Exhibit 1.7. lists the sort criteria. [10]

Typically, a single employer's activity should be counted only once among the transactions listed on the ETA 581 report. One exception is population 3, status determinations. More than one status determination may occur during the quarter covered by the report. For example, an employer newly determined as liable as a covered employer may later acquire another business, and a second status

[9] They also should be used by the programmer as part of the initial specifications in Appendix A for building the reconstruction file. Invalid duplicates are not reportable, and by eliminating obvious duplicates from the reconstruction file, the validator can focus on identifying invalid duplicates that are not obvious. Every effort should be made to eliminate invalid duplicates as the validation extract files are created.

[10] If the reported counts and the reconstruction counts are from the same data source, such as a statistical file, duplicates could impact both counts equally. If the two sets of counts are from different sources, count differences could indicate a greater number of duplicates in the file with the higher counts.

determination would be required to determine the status of that acquired business. As a result a single EAN may appear on two or more transactions. Since no duplicate *transactions* are allowed in the reconstruction files, validators will check other information about the transaction to make sure each transaction for an EAN represents a unique status determination. For example they might look at determination date or predecessor number to see whether they are required to determine the successorship status of the employer acquiring the business.

B. MATERIALS

1. Four Duplicate Detection Files Sorted by EAN

Duplicate detection procedures are used for four of the five populations (populations 1, 2, 3, and 5). There is no duplicate detection validation for receivables (population 4) because dollar amounts cannot be adjusted based on samples.[11]

As specified in Exhibit 1.7, a sample of 400 random EANs is selected for examination in each of the four populations. The sample listing contains ***all transactions in the population that have one of these 400 EANs***. (Thus the file may contain more than 400 individual transactions for review.) Once the sample files are selected, they are sorted according to the directions in Exhibit 1.7. The primary sort is by EAN. The secondary sort (when applicable) is by the second field indicated in the Data Elements column in Exhibit 1.7. This column also provides rules for the validator to determine whether duplicate transactions are valid.

For example, for population 1, the programmer should select all transactions associated with 400 unique EANs from the population 1 validation file. Because it is not practical to identify duplicates by subpopulation, ***the sample is drawn from the entire population***.

The Data Elements column on Exhibit 1.7 indicates which data elements from the RV file specifications are needed. The programmer is *not required* to include every data element from the RV file specifications on the duplicate detection file. However, it is *prudent* for States to include all data elements from the RV file. Using the full complement of data items will provide more information for analysis, while taking advantage of the file layout already created for the main RV file.

[11] However, in building the reconstruction file, programmers should check for duplicate receivables transactions by reviewing the EAN, Established Date, employer report quarter, and amount to make sure that population 4 is accurately reconstructed. There should be no duplicate transactions.

EXHIBIT 1.7

DUPLICATE DETECTION CRITERIA, BY POPULATION

Population Description	Number of EANs or Duplicates[a]	Duplicate Detection Criteria	Data Elements
1. Active Employers	400 EANs	The EAN should not appear twice. (Multiunit employers are counted as one employer.)	EAN (Column 1)
2. Report Filing	400 EANs	The EAN should not appear twice. (If an EAN has reports for multiple employer report quarters (ERQ), only the report for the ERQ immediately preceding the report quarter (RQ) is countable. Reports from multiunit employers are counted as one report.)	EAN (Column 1), ERQ (Column 2)
3. Status Determinations[b]	400 EANs	No single transaction should appear twice. Individual EANs may appear more than once. For example, there might be two transactions listed for a single EAN if an employer acquires two businesses at different times during the quarter, resulting in two successorship determinations. Multiple determinations may be legitimate, as long as they do not reflect clerical errors.	EAN (Column 1), Status Determination Type Indicator (Column 3), Status Determination Date (Column 5), Status Determination Processing Date (Column 8, 9, 10, 12, or 13), Predecessor Account Number (Column 11)
5. Field Audits	400 Audit IDs	The same Audit Identification Number should not appear twice.	EAN (Column 1), Audit ID# (Column 2)

[a] 400 EANs are randomly selected from the RIV file and scanned. The State should list all transactions for each EAN selected. Since the programmer should have incorporated duplicate detection logic into the programming of the RIV file, the expectation is that no ***invalid*** duplicates will be found.

[b] There may be issues related to duplicate detection when a statistical file contains transactions that are not present on the database. States may not be able to perform this sort.

[c] If the State identifies a unique identifier for each determination, the unique identifier should not appear twice, and the population is validated for duplicates by the same method as populations 1 and 2.

2. Four RV Spreadsheets

The spreadsheets for populations 1, 2, 3, and 5 have space where validators can report the results of the duplicate detection process. The spreadsheets adjust the overall validation counts as appropriate when duplicates are found. (See Module 1.1.)

C. METHODOLOGY

1. Select a sample of 400 EANs for populations 1, 2, 3, and 5. Draw the sample from the entire population rather than individual subpopulations. In the list of transactions for the sample, include all transactions associated with each EAN in the sample. There may be more than 400 transactions on the list. Sort the list of sampled transactions by EAN and perform secondary sorts as specified in Exhibit 1.7.

2. Examine the sampled transactions, reviewing source documentation as necessary to assure that none of the transactions are duplicates.

3. Record the number of ***transactions reviewed*** and the ***# of duplicates***, found in the ***Duplicate Calculations*** box at the bottom of the RV spreadsheet. The spreadsheet will automatically calculate the percentage of duplicates and enter that percent in the Total row of the ***% Duplicate Errors*** column. For example, for active employers, 44 duplicates found in a sample of 440 checked would equal a duplicate errors percentage of 10 percent.

   Count every duplicate. For example, if a single transaction is listed twice, then count one duplicate. If the transaction appears three times, then there are two duplicates.

4. The RV spreadsheet will automatically update the ***Adjusted Validation Count*** in the Total row after the % Duplicate Errors has been calculated. It will do so by reducing the ***Validation Count*** by the percentage of duplicates found in the file.

5. The RV spreadsheet calculates the ***Count Difference*** and the ***Count % Difference*** *for the total population* after adjusting for the percent of Duplicate Errors in the sample.

D. EXAMPLE

Exhibit 1.8 is an example of a file for the population of active employers that has been sorted by EAN.[12] The validator scans the EANs for employers that appear twice. (Since the programmer should refer to the duplicate detection criteria in building the reconstruction file, the expectation is that few, if any, invalid duplicates will appear.) In this example, there are no duplicate transactions among the eight employers. Thus in the Duplicates Checked box on the RV spreadsheet, the validator enters 8 for transactions reviewed and 0 for number of duplicates. The spreadsheet divides 0 by 8 and enters 0% in the % Duplicate Errors column on the Total row of the RV spreadsheet.

Had one of the transactions been a duplicate, the State would have entered 8 transactions checked and 1 for the number of duplicates. The spreadsheet would have calculated a % duplicates of 12.5% (1/8).

E. RESULTS

The **% Duplicate Errors** will be automatically calculated on the Total row of column six of the RV spreadsheet. This value is used to calculate an adjusted validation count, which, in turn, is used to calculate the error rate in the count for each population. Invalid duplicate transactions should be researched first by population for systemic problems, such as a one-day file corruption, and then by transaction. Errors should be brought to the attention of the programmer and noted in the Comments column of the RV spreadsheet.

[12] The example file for active employers only contains eight transactions (employers) to illustrate duplicate detection. In reality, the validator would draw a sample of 400 EANs (about eight printout pages). If there are duplicates then some EANs may appear more than once, creating a list of more than 400 individual transactions.

EXHIBIT 1.8

ACTIVE EMPLOYERS SORTED BY EAN

Obs.	1 Employer Account # (EAN)	2 Employer Status Indicator A/I/T	3 Employer Type C/R	4 Liability Date (Reopen)	5 Liability Date (Initial)	6 Inactive/ Terminated "as of" Date	7 Activation Processing Date	8 Number of Liable Quarters	9 Sum of Wages (Past 8 Q's)
5	123	Active	C		04/01/95			8	950,000
4	234	Active	C		01/01/94			4	58,000
6	235	Active	C		10/01/95			8	100,000
1	567	Active	C		07/01/37			8	827,000
3	687	Active	C		01/01/90			2	23,000
8	747	Active	R		10/31/95			8	832,000
2	891	Active	C	3/31/94	01/01/52	09/30/93		8	400,000
7	898	Active	R		04/01/92			7	62,000
Errors	0								

F. ALTERNATIVE METHODOLOGY

Frequency counts generated by standard statistical software packages can be used to test for duplicates. By running frequency counts of EANs, for example, EANs that appear more than once can be seen quickly. It will still be necessary to examine duplicates that are identified to determine if they are true duplicates and their cause.

Appendix F describes this alternative methodology.

G. OVERVIEW

Exhibit 1.9 summarizes the key tasks involved in duplicate detection and suggests staff roles for each task.

EXHIBIT 1.9

TASK 3: DUPLICATE DETECTION

Activity	Roles
For populations 1, 2, 3, 5 select a random sample of 400 employer account numbers. Include every transaction in the population that is associated with each EAN in the sample.	ADP, Validators
Print out the records for employers in the sample after sorting them by EAN and by secondary sort criteria indicated in Exhibit 1.8.	ADP
Review any duplicate listings to ensure that they are true duplicates. Use the duplicate detection criteria in Exhibit 1.8 to guide the review.	Validators
Enter the duplicate detection findings in the RV spreadsheet.	Validators

Module 2

TRANSACTION VALIDATION

Before validators use reconstruction files to validate reported counts, they must assure themselves that those files are as accurate as possible. The process of testing whether the reconstruction files are accurate began with Module 1.2, Duplicate Detection. That process makes sure that each transaction is counted only once. It searches out invalid duplicate transactions whether those duplicates are produced by the reconstruction process, or whether they already existed in the State's UI employer database.

Module 2 focuses entirely on the elimination of errors that occur as the reconstruction files are built. The reconstruction validation (RV) files are extract files. They are built from information taken from the State database.

Module 2.1, File Integrity Validation (FIV), checks that the correct information was brought over from the database to build the reconstruction file.

Module 2.2, Data Element Validation (DEV), checks that transactions were assigned to the correct subpopulations, following the specifications contained in this handbook.

Once these checks are complete, and any problems resolved, the State validation team can be confident that the validation counts are correct. If, in the end, there are differences between the ***validation*** counts and the ***report*** counts, States will conclude that the *reported* counts are not valid.

MODULE 2.1 — FILE INTEGRITY VALIDATION (FIV)

A. PURPOSE

Module 2.1 checks whether each transaction, and each data element listed with the transaction, is an accurate representation of the information in the State UI contribution database. This check can be done using a very small number of transactions. The process that extracts information from the database and places it in the reconstruction file is highly automated. Automated processes are repetitive. If, for example, a certain field in the employer history file is extracted and placed in the fifth column of the reconstruction file for a single transaction, that same field will be used for the fifth column of every transaction. Thus, if we know that all data elements have been transferred correctly for a few transactions, we can be assured that all similar transactions are done correctly.

To conduct File Integrity Validation, the validator compares all characteristics listed for two transactions that have been reconstructed to all available supporting documentation in the State's database. By checking the data in every column the validator ensures that the data in the reconstruction file are accurate.

Elements requiring data from multiple fields pose a greater risk of reconstruction error. For example, the reactivation date for status determinations may not come directly from one field in a State's database, but instead from a combination of a transaction code and a transaction date field. There may be a series of applicable transaction codes representing reactivations. In these instances, the State or region may want to examine these elements in greater detail.

B. MATERIALS

1. RV Files and Printouts

The validators select two cases for review in each subpopulation. They can do so most easily by turning to the first page of the reconstructed report validation (RV) file for each subpopulation (created in Module 1.1). There they select the first two cases (i.e. the first two rows of the file). They mark the chosen transactions and retain that page of the printout. This creates an audit trail of the process, as well as a hard copy for review by the Federal regional auditor.[1]

2. Query Screens

As they review the data items in each transaction, validators refer to query screens from the State data system. These screens display information on transactions and the status of employer accounts. It is strongly recommended that the necessary screens be printed at the time the reconstruction file is originally created. This will reduce the chance that the printout and the query screen differ only because time has passed and data haves changed between the time the screen printout was produced and the transaction file created.

3. State-Specific FIV Instructions

Module 3 of the handbook includes FIV instructions that use the State's own screen and field names. For each data element to be validated (each column of the RV

[1] Validators are also encouraged to review any transactions they see where values seem unusual. The presence of "outliers" may indicate a problem that needs correction.

specifications), the handbook lists one or more steps that guide the validator through the FIV process. Each "step" in the FIV portion of Module 3 validates a discrete transaction characteristic. The validator compares the reconstructed data element to the data on an employer history screen to ensure that the reconstruction file was accurately programmed. The validation steps document the field names and values used on the State screens. The validator must check the data element against these criteria to ensure that the data were correctly transferred from the database to the reconstruction file.

4. Five Validation Worksheets

Validation worksheets, one for each population, guide the validator to the appropriate step in the State-specific handbook where instructions are provided for the validation of each data element. The worksheets list the subpopulations to be validated under the heading File Integrity Validation. The validator also uses the validation worksheets to summarize and evaluate errors identified through this process. These worksheets can be found in Appendix E of this handbook.

C. METHODOLOGY

To conduct file integrity validation, perform the following steps.

1. Identify the subpopulations to be validated by consulting the top portion of the validation worksheets under the heading File Integrity Validation.

2. Go to the first page of the printout for each subpopulation and select the first two cases for validation.

3. Validate the selected transactions on the printout by checking each item (column) against the corresponding field on the database screens printed from the employer master file. It may be helpful to print a special version of this first page with a narrow, extra blank column next to each existing column. This extra column can be used to mark whether or not each item in the selected transaction is valid.

4. The "Step" number in the column heading of the reconstruction printout directs the validator to the appropriate page in the State-specific handbook (Module 3). There the validator will find, rules that document the required screens and the logic tests that must be done for each element.

5. Each step listed in Module 3, **File Integrity Validation Instructions,** helps the validator locate and compare specific pieces of information from the

supporting documentation with the corresponding data on the reconstruction file to determine the validity of the information (pass or fail).

The instructions for each step or substep name the supporting documentation (screens) that the validator will need. A set of logic tests, called validation *rules*, determines the accuracy of each characteristic of a given transaction. A subsection, called *function*, explains the purpose of each *rule*.

The instructions will indicate if the State does not maintain a specified data element, if the State maintains it but not on an existing screen, or if the State cannot validate a data element through the automated validation process.

Definitions listed within each step in Module 3 give the Federal definition of the item being validated. The Federal definition is followed by *examples*, *includes (*situations falling within the definition*)*, and *excludes*, which provide further information on the data element.

Definitional Issues describe known discrepancies between State and Federal definitions. This section serves an important role in systematically documenting validation issues in advance, letting validators and auditors know when problems are anticipated. State staff were interviewed during the design of data validation. Known issues were listed at that time and additional issues will be added by States as they identify issues during the validation process.

Comments provide additional information that State staff or Federal regional auditors may need to handle unusual situations.

6. Put a check (✓) on the reconstruction printout next to each data element after that data element successfully passes a step.

7. After completing two transactions for a subpopulation, record the total number of, transactions checked at the bottom of the printout (usually 2) and the total number of errors below each column. The validator can use the validation worksheets as a guide to further researching errors. The reconstruction file must be reprogrammed until FIV finds no errors.

8. When the integrity of the reconstruction file is successfully validated for each of the subpopulations, the validator completes the subpopulation row on the validation worksheet by entering the Number of Cases Checked and placing a check mark in the Pass box on the worksheet.

9. Save the printout with checkmarks next to the validated items, the Validation Worksheet and any screen prints. They will be used during the Federal review of the validation process.

D. EXAMPLE

Exhibit 2.1 contains a sample validation file for population 1, active employers, and a sample of the completed FIV portion of the corresponding population 1 validation worksheet. The validation worksheet instructs the reviewer to validate the data elements in every column in the file for two transactions per subpopulation. (The corresponding steps for these columns are indicated in the column headers in the file.)

The validator, beginning with the first column on the left of the file, refers to the indicated step number in the State-specific portion of the handbook and follows the FIV instructions. The number of errors identified for that data element for the minimum sample is written on the printout beneath that column. (The validator may circle incorrect values in each transaction before summing the values in error for the sample.)

In this example, the validator found no errors in FIV and indicated a "Pass" on the validation worksheet.

E. RESULTS

If the data in the reconstruction file do not match the data on the screen, the reconstruction file must be reprogrammed until the FIV reveals no errors. The reconstruction file is the basis of all validation exercises and must be proved valid before proceeding any further.

EXHIBIT 2.1

VALIDATION FILE—ACTIVE EMPLOYERS—POPULATION 1

Obs.	1 *(Step 1A)* Employer Account # (EAN)	2 *(Step 3A)* Employer Status Indicator A/I/T	3 *(Step 2)* Employer Type C/R	4 *(Step 4B)* Liability Date (Reopen)	5 *(Step 4A)* Liability Date (Initial)	6 *(Step 5)* Inactive/ Terminated "as of" Date	7 *(Step 15)* Activation Processing Date	8 *(Step 7B)* Number of Liable Quarters	9 *(Step 7A)* Sum of Wages (Past 8 Q's)
1	567	Active	C		07/01/37			8	827,000
2	891	Active	C	3/31/94	01/01/52	9/30/93		8	400,000
3	687	Active	C		01/01/90			2	23,000
4	234	Active	C		01/01/94			4	58,000
5	123	Active	C		04/01/95			8	950,000
6	235	Active	C		10/01/95			8	100,000
Subtotal	6								
1	898	Active	R		04/01/92			7	62,000
2	747	Active	R		10/31/95			8	832,000
Subtotal	2								

VALIDATION WORKSHEET
Active Employers

File Integrity Validation				
Step Type	**Subpopulations**	**Column #**	**Number of Cases Checked**	**Pass/Fail**
FIV	1.1	All	2	Pass
FIV	1.2	All	2	Pass

F. OVERVIEW OF MODULE 2.1

Module 2.1 provides tools to test that the data used to create RV files accurately reflects the information in the State's UI database. Exhibit 2.2 summarizes the steps in the File Integrity Validation Process.

EXHIBIT 2.2

TASK 3: FILE INTEGRITY VALIDATION

Activity	Roles
Secure a printout of the first page of the RV printout for each subpopulation. Mark the first two records.	Validator, ADP
In advance, have necessary query screens produced at same time reconstruction file is created.	Validator or ADP
Following the Steps indicated in the Tables in Appendix A and described in Module 3, review and validate every item (column) on the printout for the two selected records. The review compares information listed in the reconstruction file to source documentation, typically query screens on the UI database.	Validator
Record the results on the top half of the worksheet found in Appendix E**.**	Validator
If, invalid data were used in the creation of the reconstruction file, correct the file and begin this task over again. This is obviously a step that should be done well before the scheduled date for the actual validation.	ADP
Conduct Item Validation for corrected versions of the RV file.	Validator

MODULE 2.2 — DATA ELEMENT VALIDATION (DEV)

A. PURPOSE

One of the most important goals of the data validation process is to ensure that the individual transactions reported on ETA 581 are classified correctly. Data Element Validation (DEV) tests that all cases listed on each validation file are, indeed, located in the appropriate subpopulation. After completing Module 2.1 to ensure that the correct data elements are used in listing transactions, the validator, goes on to test that all transactions meet the definitional parameters for inclusion in a reporting category (subpopulation). In theory, this may be accomplished by visually scanning an entire printout to ensure that the patterns in the data are correct (for example, all reimbursable employers have an account number greater than 90,000 or an indicator of "R").

While this method may be satisfactory when the files are small and data are arranged in easily visible patterns, it is less practical when using the very large files reconstructed for data validation. In addition, correct classification of employer contribution data often involves relationships between dates, and these relationships do not lend themselves to easy visual scanning by the validator.

Therefore, for Data Element Validation, the data are ***sorted*** in different ways to enable the validator to easily detect invalid data. After the file is sorted, all "out of range" values will appear at the beginning or end of the file.

Sorting requires minimal additional programming. Programmers use the exact same files and file formats created for the original reconstruction file (see Module 1.1 and Appendix A). For sorting, they are only asked to present the transactions on the file in different orders. Once the data are sorted, the validators can quickly and efficiently determine if all subpopulations were reconstructed properly. Overall, sorting the file is more efficient and thorough than manual review.

There are two types of sort specifications in Appendix A. Those with an "S" as the second character (e.g., sort 1.S.2) are done during the building of the original reconstruction file. These sorts simply restate the Sort Keys listed as part of the specifications in Appendix A, Part I. They do not require any additional printouts. However, validators need to check the original RV printout to ensure it conforms to the associated sort criteria.

The second type of sort specification has a letter other than S in it (e.g. sort 2.A). These sorts do require ADP to create separate versions of the reconstruction file, each in a different order.

There are multiple sorts because the inclusion of transactions in a subpopulation is often conditioned by the date on which the transaction occurred. For example, we often need to ask the question, "Did the transaction occur in the report quarter?" If the file is sorted by transaction date in ascending order, then this question is easy to answer. Any transaction that occurred too early, i.e., before the report quarter, will appear at the beginning of the list. Any transactions that occurred too late, i.e.,after the end of the report quarter, will slide to the end of the list.

It is sometimes necessary to check more than one date. This requires a second sort if the file was previously sorted by date. As soon as the file is sorted by the second date, it is no longer in order by the first date. Thus, a separate sort needs to be done for each date to be checked.

B. MATERIALS

1. RV Files

See Module 1 and Appendix A for a detailed explanation including the criteria for organizing each transaction population into its subpopulations.

2. Sorted RV Files

Several additional versions of the RV files are produced. They are identical in layout and content to the original RV files. However, each version is sorted in a different order as specified in the data element validation specifications.

3. Data Element Validation Specifications

Appendix A, Part II provides a detailed explanation of the reconstruction file sorts and the subsequent sorts for each of the five populations. The sort criteria help ADP staff understand the programming logic used to generate the reconstruction and sort files.

The State must print the first and last page of each sort from the reconstruction file and from the sort files.[2] This creates an audit trail and enables the Federal regional validator to review the materials off-site.

[2] Printing the first and last page of the file is usually all that is necessary. Only a few transactions, if any, will fall outside the allowable range. However, if there is a major programming error, several pages at the beginning or end of the file may contain invalid transactions.

4. Validation Worksheets

The five validation worksheets guide the validator through the Data Element Validation as well as the File Integrity Validation. Under the heading DEV, the worksheets indicate the sorted files to be validated and the relevant steps and rules used to validate each. The validator also uses the validation worksheets (found in Appendix E) to summarize and evaluate errors identified through the DEV process.

5. State-Specific TV Instructions

Each step in the DEV section of the TV instructions validates a discrete transaction characteristic. Many validation steps are divided into substeps. The steps and substeps are the lowest common denominators of the reportable UI employer contribution transactions being validated; they combine to create the distinct subpopulations on the RV files.

Each data element sort is included in the appropriate step or substep for the particular element being validated. For example, in Step 3.A, Active Employers, there are four sort rules.

C. METHODOLOGY

1. Sort all observations within or across each subpopulation by the reconstruction file sort keys. (This file is also used in Module 1.)

2. Follow the instructions in Appendix A, Part II, to produce additional sorted versions of the RV files. (Each sort should include the full table header indicated on the RV file specification and should be saved as a separate file.)

   If the Data Element Validation indicates that the files have been built incorrectly, the programming should be corrected and re-run. To allow time to correct problems, it is wise to conduct test runs and make corrections well before the actual validation is scheduled.

3. Print the first and last page of each sorted version of the reconstruction file.

4. The validation worksheets (found in Appendix E) guide the validator to the appropriate steps in the State-specific portion of the handbook (Module 3).

5. When referred to a designated step or substep by the validation worksheet, the validator follows the sort rules provided.

6. For each step, **Data Element Validation Instructions** tell the validator to scan the beginning and/or the end of the sort for adherence to specified criteria (e.g., dates must be in a certain range or codes have certain values). In Module 3, these instructions appear below the FIV instructions (described in Module 2.1) for each step. The instructions will indicate if a sort is not applicable to the State, because the sorted data element is not maintained or for another reason.

7. For sorts 1.C and 3.S.5, the validator must examine a few transactions to supplement the sort (see the sort specifications for details). Sorts test whether the data in the reconstruction file meet the parameters of the ETA 581 report. However, for a few data elements, the validator must take additional steps to test the reconstruction.

   Sort 1.C requires the validator to examine a printout of 200 employers who have an inactivation date but are reported as active. The review ensures that the inactivation date is prior to a reactivation date. If the validator examines a sample of 200 employers with inactivation dates and determines that in each case there was a reactivation date after the inactivation date and by the end of the report quarter, it will verify that the State is reporting correctly.

   Validation of sort 3.S.5 requires a review of a sample of five outliers (cases with extreme values) to verify that there are no reporting errors in the extreme cases of this file.

8. On the validation worksheet, the validator records the total number of transactions checked and the number of errors found. The total number checked is usually the subpopulation count.

9. The percentage of errors is computed by dividing the number of errors by the total number of transactions checked.

10. If the percentage of errors is less than 2 percent, the validator records that the sort passed. If the percentage of errors, is more than 2 percent, the validator records that the sort failed.[3]

11. The validator also records the counts from each sort in the Sort Count column on the RV spreadsheet, which computes differences between the sort counts and the corresponding validation counts.

[3] Acceptable error levels to be confirmed by the U.S. Department of Labor National Office.

12. Retain the first and last page of each sort and the Validation Worksheets. These materials will be used by Federal reviewers.

D. EXAMPLE

Exhibit 2.3 shows active employers sorted by the number of consecutive quarters of liability for the employer, up to eight, in descending order. It also sorts the employers by the sum of the wages from contribution reports (or delinquent quarters or assessments) for those quarters in ascending order. This sort ensures that no employers will be counted as active if they have filed contribution reports for eight consecutive quarters and each of those reports has zero wages listed. After the sort is generated, the validator follows the rules for validating this sort as they appear in the handbook.

Continuing with this example, sort 1.D on the validation worksheet directs the validator to Step 3.A, Rule 4 (see example). This rule instructs the validator to check the listing of the sum of the wages, to see if the first observation has a zero sum of wages for those EANs reporting wages for at least eight consecutive quarters. If the sum of wages is greater than zero, the validator knows that all observations meet the requirements. The validator then puts a check mark under the applicable column and completes the appropriate boxes for that sort on the validation worksheet.

If the first sum is zero for eight quarters, the validator scans the listing in order to see how many other observations have a sum of zero wages for eight quarters and are thus "invalid" transactions. The number of errors is written under the applicable column. The validator then researches the invalid transactions to determine why they were included in the validation reconstruction and how they might impact the count comparison between the reconstruction and the ETA 581.

In this example there are no values out of range so the total number of cases checked would be entered in the sort count column of the RV spreadsheet next to the appropriate sort number.

EXHIBIT 2.3

ACTIVE EMPLOYERS SORTED BY NUMBER OF LIABLE QUARTERS AND SUM OF WAGES

Obs.	1 *(Step 1A)* Employer Account # (EAN)	2 *(Step 3A)* Employer Status Indicator A/I/T	3 *(Step 2A)* *(Step 2B)* Employer Type C/R	4 *(Step 4B)* Liability Date (Reopen)	5 *(Step 4A)* Liability Date (Initial)	6 *(Step 5)* Inactive/ Terminated "as of" Date	7 *(Step 15)* Activation Processing Date	8 *(Step 7B)* Number of Liable Quarters	9 *(Step 7A)* Sum of Wages (Past 8 Q's)
6	235	Active	C		10/01/95			8	100,000
7	90452	Active	R		9/4/51			8	347,099
2	891	Active	C	03/31/94	01/01/52	09/30/93		8	400,000
1	567	Active	C		07/01/37			8	827,000
5	123	Active	C		04/01/95			8	950,000
8	90211	Active	R		01/01/67			8	21,444,000
4	234	Active	C		01/01/94			4	58,000
3	687	Active	C		01/01/90			2	23,000
Correct?									

VALIDATION WORKSHEET 1
Active Employers

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**Number of Cases Checked**	**Number of Errors**	**Percent of Errors**	**Pass/ Fail**
Sort	1.S.1	Employer status	3A/1				
Sort	1.S.2	Employer type	2A/1				
Sort	1.S.3	Employer type	2B/1				
Sort	1.S.4	1. Liability date (reopen) 2. Liability date (initial)	3A/2				
Sort	1.A	EAN	2A/2				
Sort	1.B	EAN	2B/2				
Sort	1.C	Inactive/terminated date	3A/3				
Sort	1.D	1. Number of liable quarters 2. Sum of wages	3A/4	8	0	0	Pass

E. RESULTS

The validator records the total count for each sort on the RV spreadsheet and records any invalid transactions identified for each data element on the validation worksheet. Error rates are calculated for each sort. If the review reveals significant errors in the construction of the report validation files, the programming must be corrected and the process repeated.

At this point the validators have completed the three tests of the accuracy of the reconstruction files:

- Duplicate detection,
- File Integrity Validation to assure the correct data have been selected from the UI database, and
- Data Element Validation to assure all transactions are assigned to the correct subpopulation.

Where necessary the data validation files have been fixed and results of the reviews have been documented. Validators now return to the RV spreadsheets and enter the final validation counts. If the results indicate that information on unemployment insurance reports is incorrect, States will be asked to undertake corrective action.

F. CORRECTIVE ACTION

Validation is not an end in itself; it is a means toward correct reporting. If validation identifies reporting errors, the State should correct them as soon as possible, even if their magnitude is below validation limits.

Corrective Action Plan. To document the actions required for corrective action and the timetable for their completion, the State must provide its their ETA Regional Office a brief Corrective Action Plan (CAP) containing the following information on every validated report element found to be in error by more than validation limits:

- *Report element(s) in error*
- Magnitude of error found
- *Status/Plan/Schedule for correcting* (if, as happened in the validation pilot, reporting errors were corrected in the course of the first validation, the report

should simply note "corrected during validation"). The plan may be multi-year.

- *Procedures used to re-validate items corrected during validation*.

Timing of CAP. The CAP must be submitted within one month of submitting the State's Validation Summary Report. CAPs are considered additions to the State's Quality Service Plan (SQSP). If the State is conducting the validation in segments, e.g., Benefits first, then Tax, and a CAP is required based on a segment's validation results, it must be prepared within a month of the completion of the segment.

Revalidation. Every element in error by more than the stated validation limit must be revalidated the following year. A "clean" validation confirms the success of the corrective action taken or, if the State has not completed corrective action, identifies the current extent of error.

Errors Discovered Outside the Validation Process.. Such errors are considered serious because they were not detected by State validation staff and validation procedures. Such errors require either a new CAP or — if the State has an active CAP because of a previous validation — an amended CAP.

G. OVERVIEW OF MODULE 2.2

Module 2.2, Data Element Validation, has tested whether transactions reported on the UI Contribution Operations report have been placed in the correct category. Exhibit 2.4 summarizes key tasks the State must go through to complete DEV. Once the DEV test is passed, the final version of the validation findings can be entered on the RV spreadsheets. Exhibit 2.5 summarizes those entry tasks.

EXHIBIT 2.4

TASK 4: DATA ELEMENT VALIDATION
SORTING AND REVIEWING THE RECONSTRUCTION FILES

Activity	Roles
Identify sort requirements from Appendix A, Part II.	ADP, Validators
Sort RV files as indicated in the Appendix. This will produce several new files for each subpopulation. Retain the same headings and record layout as in original file. Change only the order of transactions and the title of the printout.	ADP
Review the first test of sorted files. Make any revisions needed to the reconstruction file.	ADP, Validator
Print out and review the first and last page of each sort file. The review should follow the Steps listed in the sort criteria, and discussed in the Data Element Validation instructions in Module 3.	Validators
Review additional transactions for sorts 1.C and 3.S.5.	Validators
Return to the RV spreadsheet and enter the final data validation findings.	Validators

EXHIBIT 2.5

TASK 5: RECORDING THE RESULTS

Activity	Roles
Enter report counts and validation counts in the spreadsheets found in Appendix B. This will normally be done manually, although States have the option to develop processes to automatically transfer information into spreadsheets.	Validator or ADP
The spreadsheet will calculate a Pass/Fail verdict on the accuracy of each validated item.	
Submit the findings as documented on the spreadsheets.	Validators

Module 4

TAX PERFORMANCE SYSTEM
ACCEPTANCE SAMPLE VALIDATION

A. PURPOSE

Tax Performance System (TPS) Acceptance Sample Validation reviews sample selection procedures used by TPS (formerly RQC, Revenue Quality Control). It ensures that the samples used to assess status determination and field audit quality are randomly selected from the correct populations.

There are two basic approaches to selecting samples. The first approach is a conventional interval sample. Here the programmer (or a utility program) divides the size of the desired sample (say 60) into the size of the population (say 600) and derives the sample interval (every 10th observation). The programmer or the utility program then selects a random start point (in this instance) between 1 and 10 and selects every tenth case from that point.

The second approach is to use a sampling utility program that randomizes the file and selects the first 60 observations. This approach is somewhat more difficult to validate, but could involve a review of the sample against the source file or review of the utility program specifications.

B. MATERIALS

1. Copy of the Universe File of Status Determinations for the Quarter

One for each of the following TPS universes:

- New
- Successor
- Inactive/Terminated

2. Copy of the Universe File of Field Audits for the Quarter Used for the TPS Sample

C. METHODOLOGY

1. Compare the total count for the quarter of the three status determination universes and one field audit universe to the count reported on the ETA 581 for the same period. This validates that the correct universe was used.

2. Determine if an interval sample was drawn (and how it was drawn) or if the file was randomized such that the first set of cases could be selected without establishing intervals.

3. If an interval sample was drawn, check to see that the proper cases were selected (that is, if the random start was 10 and the interval was every 40th case, check to see that cases 50, 90, 130, and so forth were selected). The validator can identify the sampled cases from the quality review documentation.

4. If the sample was drawn from a randomized file, print the file, ensure that it was not ordered by date, employer, or some other nonrandom means. The validator can compare the printout with the way the file was ordered prior to randomization to ensure that the file was randomly reordered.

D. RESULTS

If the sampling method was not correct or was not implemented properly, the validator should discuss the problems with the programmer. If the programmer confirms that the process was incorrect, the validator should record the problems on the validation worksheet, included in Appendix E, for the two samples.

Module 5

WAGE ITEM VALIDATION

A. PURPOSE

Wage Item Validation verifies that the wage item transactions processed in the report quarter are accurately reported on the ETA 581. This helps ensure equitable funding when this item is used to determine State workload. A wage record is the listing of an individual's earnings in covered employment. Each individual employee's earnings are listed by Social Security Number (SSN) and are submitted by employers each quarter. Employers may submit wage records as paper records, submit them electronically, or provide, computerized files stored on magnetic tapes, diskettes, or CD-ROMs.

Validation addresses the reporting system's accuracy by verifying that the following items are not reported:

- Corrections (the system must be able to process corrections without double-counting the item).
- Incomplete wage records (for example, if the identifier or wage amount is missing for the employee).
- Duplicate records.

B. MATERIALS

1. Sample Batches of Each Mode of Wage Items for a Given Time Period

A State can use a variety of methods including magnetic tape, punched cards, key entry, diskettes, CD-ROMs, and electronic transfers.

2. Wage Item Validation Worksheet

See Appendix E.

C. METHODOLOGY

1. Identify the specific modes of data capture used for processing wage items on the Wage Item Validation Worksheet (see Appendix E).

2. Select representative periods of time when each of these modes is in use. (This time period should be indicated on the worksheet for each of the applicable modes.) It may be necessary to select batches from different days to ensure that each mode of data capture is examined.

   Normally states will be expected to select and review at least five batches of wage reports for wage items that have been entered manually. For information transmitted electronically, the validators should select at least a day's worth of entry.

3. For each of the applicable modes on the Wage Item Validation Worksheet, the validator enters the number of wage items reported in the ETA 581 count for the particular batch being examined. This information must be obtained from the system used to compile the wage item count for the ETA 581.

4. Recount the number of wage items in each of the batches, for each mode, using the Federal definition for a countable wage item.

5. Ensure that there are no duplicate entries — that each wage record is counted only once.

6. Corrected wage items are counted only if they were not previously included.

7. The validator must count only wage items that are complete. This means each processed entry should include the following information:

   - Employee Identifier (Name or SSN)
   - Employer Identifier (Name or EAN)
   - Wage dollar amount

   If a wage record is incomplete, count only those records containing a dollar amount and another element that positively identifies the worker either by name or SSN or by employer name or account number.

8. Enter the total number of wage items included in the recount on the **Wage Item Validation Worksheet**. If any duplicates or errors have been identified, the validator indicates these errors in the appropriate columns on the worksheet.

D. EXAMPLE

Figure 5.1 shows an example of a Wage Item Validation Worksheet listing a number of possible modes of wage item processing in the first column. In this particular State, the validator has chosen only the applicable modes and selected an appropriate time period for each mode. The column labeled "581 Count for Batch" has been filled in with the number of wage items processed in this batch as reported on the ETA 581. Once the validator has recounted the wage items for each of the modes, this number is reported in the column labeled "Recount for Batch."

In this example, the wage items that were electronically transferred and processed using CD-ROMs show no discrepancies between the two counts and are therefore proven to be valid. The recount of the magnetic tape processing, however, indicates a duplicate wage item, and therefore the counts do not match. This requires further research to establish the reason for the miscount and to correct any other errors caused by the use of this mode of processing.

E. RESULTS

If the wage item processing method is proved invalid through the recount process, the validator should discuss the problems with the programmer or individual responsible for wage item processing, and the necessary efforts should be made to determine if the error may affect other batches of wage items as well.

EXHIBIT 5.1

WAGE ITEM VALIDATION WORKSHEET

Mode	Time Period	581 Count for Batch	Recount for Batch	Number of Errors	Percentage of Errors	Missing ID	Missing Amount	Double Count
Magnetic Tape	12/02/96 to 12/03/96	49	48	1	2.10%	0	0	1
Electronic Transfers	12/02/96 to 12/03/96	206	206	0	0.00%	0	0	0
Data Entry	12/02/96 to 12/03/96	400	392	8	2.00%	4	2	2
Scanning	12/02/96 to 12/03/96	600	600	0	0.00%	0	0	0

Appendix A

PART I

REPORT ITEM VALIDATION SPECIFICATIONS

INTRODUCTION

As described in Module 1 of the handbook, the first step in the data validation process is to create report validation (RV) files (also referred to as extract or reconstruction files). These files list all transactions of a single type that are to be reported on the ETA 581 report. Each transaction is listed in a single population and in only one subpopulation within the population. This appendix provides the detailed specifications for creating the RV files (Part I) as well as the criteria for creating versions of the file sorted in different orders (Part II).

Tables 1 through 5 are reconstruction file specifications for each of the validation populations described in this handbook. Each row of the table is the specification for a single, mutually exclusive subpopulation. State programmers should create reconstruction files by following the specifications, including the specifications for table headers. Observations within each file should be numbered, and a total of the number of observations in each subpopulation should be included. The sort key row on each specification indicates the manner in which transactions should be sorted when they are listed on the reconstruction file.

At the end of each table is a verbal description of each subpopulation. This should help readers orient themselves to the information in the table.

Each column header includes a step number which refers to the State-specific portion of the handbook in Module 3. Validators and programmers should refer to the indicated step number for detailed instructions on how to validate the data in that column, as well as for the definition of the data element. Each specification includes a column and/or row entitled "ETA 581 Item #'s" which indicates the Item number on the ETA 581 that the count or dollar amount in the column or row is compared with on the RV spreadsheet.

States should reconstruct each population as specified for a recent ETA 581 report quarter (RQ). In addition, States (such as California and Oregon) that administer unemployment insurance together with other taxes should capture tax type, to

distinguish between the taxes being validated on the ETA 581 and others which are not countable on the report.[1]

Abbreviations:

RQ	ETA 581 report quarter
ERQ	Employer Report Quarter (quarter covered by employer's contribution report)
FDRQ	First day of the report quarter
LDRQ	Last day of the report quarter
FD(RQ+1)	First day of the quarter after the report quarter
FD(RQ-1)	First day of the quarter before the report quarter

[1] Some States may have other unique types of data elements which should be captured in the reconstruction file to facilitate validation. For example, some States may have an indicator for seasonal employers which would be helpful in validating subpopulations 2.7 and 2.15 on Table 2.

Table 1 Draft Reconstruction File Specifications
Active Employers

Report Quarter:
Validator:
Date:

These subpopulations constitute the unique subgroups of all active employers on the last day of the Report Quarter (RQ) covered by the ETA 581. Reconstruction should be done at the end of the RQ being validated (when the ETA 581 report program is run).

			1st Sort Key	2nd Sort Key	3rd Sort Key	4th Sort Key				
Subpopulation	Reported in 581 Item #'s	**1** *(Step 1A)* Employer Account # (EAN)	**2** *(Step 3A)* Employer Status Indicator A/I/T	**3** *(Step 2A)* *(Step 2B)* Employer Type C/R	**4** *(Step 4B)* Liability Date (Initial)	**5** *(Step 4A)* Liability Date (Reopen)	**6** *(Step 5)* Inactive/ Terminated "as of" Date	**7** *(Step 15)* Activation Processing Date	**8** *(Step 7B)* Number of Liable Quarters	**9** *(Step 7A)* Sum of Wages (Last 8 Q's)
1.1	1		A	C	<=RQ	<=RQ	>RQ, or <liability date (reopen), or none			(If col. 8≥8) >$0
1.2	2		A	R	<=RQ	<=RQ	>RQ, or <liability date (reopen), or none			(If col. 8≥8) >$0

Notes:

1) The 3rd and 4th sort keys should be treated as a single sort. Sort by the most recent date.
2. Either column 4 or 5 must be a date earlier than the end of the report quarter.

Subpopulation descriptions:

1.1 Active contributory employers liable by the end of the report quarter.
1.2 Active reimbursable employers liable by the end of the report quarter.

Table 2 Draft Reconstruction File Specifications
Report Filing

Report Quarter (ETA 581):
Validator:
Date:
Employer Report Due Date: The Due Date (DD) is defined as the date after which the State can impose penalty and/or interest, whichever is first applicable. It is a State-specific date.

These subpopulations constitute the unique subgroups of all employers owing contributions or required reports during the same ETA 581 report quarter (RQ), which were received timely or secured in the RQ, or reported as resolved in RQ+1. Programmers and validators should note that timely, secured, and resolved are here defined as discrete filing statuses, whereas the ETA 581 reports cumulative counts for these categories. Because of the static nature of the received date, which is a key data element for subpopulations 2.1, 2.2, 2.9, and 2.10, the entire population extract can be run at the end of RQ+1. The validation counts in subpopulations 2.1, 2.2, 2.9, and 2.10 are compared with ETA 581 counts for the RQ; all subpopulation validation counts are compared with reported counts for RQ+1 (see 581 item # references below.)

				1st Sort Key	2nd Sort Key	3rd Sort Key					
Subpopulation	Reported in 581 Item #'s	**1** *(Step 1B)* Employer Account # (EAN)	**2** *(Step 1B)* Employer Report Q (ERQ)	**3** *(Step 2A)* *(Step 2B)* Employer Type C/R	**4** *(Step 8A)* *(Step 8B)* *(Step 8C)* Filing Status Indicator (T/S/R)	**5** *(Step 9)* Received Date	**6** *(Step 10)* Final Assessment Date	**7** *(Step 4A)* *(Step 4B)* Liability Date (Initial or Reopen)	**8** *(Step 14)* Liability Date (Met Threshold)	**9** *(Step 5)* Inactive/ Terminated "as of" Date	**10** *(Step 6A)* *(Step 6B)* *(Step 6C)* Inactivation/ Termination Processing Date
2.1	6,7,(8 in RQ+1)		RQ-1	C	T	<=DD	none				
2.2	7,(8 in RQ+1)		RQ-1	C	S	> DD but <=LDRQ	none				
2.3	8 in RQ+1		RQ-1	C	R	within RQ+1	none				
2.4	8 in RQ+1		RQ-1	C	R	none	<=LD(RQ+1)				
2.5	8 in RQ+1		RQ-1	C	R	none				< RQ-1	within RQ or RQ+1
2.6	8 in RQ+1		RQ-1	C	R	none			>=RQ	>RQ and >liability date, or none	

Table 2 Draft Reconstruction File Specifications
Report Filing

				1st Sort Key	2nd Sort Key	3rd Sort Key					
		1 *(Step 1B)*	**2** *(Step 1B)*	**3** *(Step 2A)* *(Step 2B)*	**4** *(Step 8A)* *(Step 8B)* *(Step 8C)*	**5** *(Step 9)*	**6** *(Step 10)*	**7** *(Step 4A)* *(Step 4B)*	**8** *(Step 14)*	**9** *(Step 5)*	**10** *(Step 6A)* *(Step 6B)* *(Step 6C)*
Subpopulation	Reported in 581 Item #'s	Employer Account # (EAN)	Employer Report Q (ERQ)	Employer Type C/R	Filing Status Indicator (T/S/R)	Received Date	Final Assessment Date	Liability Date (Initial or Reopen)	Liability Date (Met Threshold)	Inactive/ Terminated "as of" Date	Inactivation/ Termination Processing Date
2.7	8 in RQ+1		RQ-1	C	R	none				(RQ-1)	
2.8	8 in RQ+1		RQ-1	C	R	none		=col 9 date		=col 7 date	<=LDRQ+1
2.9	9,10,(11 in RQ+1)		RQ-1	R	T	<=DD	none				
2.10	10,(11 in RQ+1)		RQ-1	R	S	> DD but <=LDRQ	none				
2.11	11in RQ+1		RQ-1	R	R	within RQ+1	none				
2.12	11in RQ+1		RQ-1	R	R	none	<=LDRQ+1				
2.13	11in RQ+1		RQ-1	R	R	none				<RQ-1	within RQ or RQ+1
2.14	11in RQ+1		RQ-1	R	R	none			>=RQ	>RQ and >liability date, or none	
2.15	11in RQ+1		RQ-1	R	R	none				(RQ-1)	
2.16	11in RQ+1		RQ-1	R	R	none		=col 9 date		=col 7 date	<=LDRQ+1

Table 2 Draft Reconstruction File Specifications
Report Filing

Notes:

1) A few States resolve reports for seasonal employers by suspending the report filing requirement in off seasons (subpopulations 2.7 and 2.15). These States should program an additional column--suspended as of quarter--with the same step reference indicated for column 9 and the parameter indicated in parentheses under column 9 for subpopulations 2.7 and 2.15.

2) States may identify all contributory and reimbursing employers who were subject to file a required report covering the quarter prior to the ETA 581 report quarter, on the last day of the quarter prior to the ETA 581 report quarter. That data file can then be used in the validation reconstruction, even though not every report owed will be resolved. (If this approach is workable for States, it can also be done every quarter to program the ETA 581.)

3) Some States may use a delinquency flag instead of the preferred received date; this creates audit trail issues to be reviewed on a State-specific basis.

4) If an employer has more than one resolved date under columns 5, 6, 9, or 10, the earliest date is considered to be the resolved date (although if possible, all resolved dates should be included on the file).

Subpopulation descriptions:

2.1 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, who filed contribution reports timely during the report quarter.

2.2 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, who filed untimely contribution reports by the end of the report quarter (secured, but not timely).

2.3 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, who filed contribution reports during the quarter after the report quarter (resolved, neither secured nor timely).

2.4 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, who did not submit a report but received a final assessment by the end of the quarter after the report quarter (resolved, neither secured nor timely).

2.5 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, who were made inactive during the report quarter, or during the quarter after the report quarter (resolved, neither secured nor timely).

2.6 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, whose liability date (met threshold) was changed from prior to the report quarter, to during or after the report quarter (resolved, neither secured nor timely).

Table 2 Draft Reconstruction File Specifications
Report Filing

2.7 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, who were suspended from filing contribution reports due in the report quarter by virtue of being seasonal employers, an administrative decision not to pursue report filing, or for other reasons (resolved, neither secured nor timely).

2.8 Contributory employers owing contributions reports for activities in the quarter prior to the report quarter, whose accounts were withdrawn by making the liability date and the inactive/terminated "as of" date equal (resolved, neither secured nor timely). This includes canceled, withdrawn, closed, dropped, etc. accounts.

2.9 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, who filed required reports timely during the report quarter.

2.10 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, who filed untimely required reports by the end of the report quarter (secured, but not timely).

2.11 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, who filed required reports during the quarter after the report quarter (resolved, neither secured nor timely).

2.12 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, who did not submit a report but received a final assessment by the end of the quarter after the report quarter (resolved, neither secured nor timely).

2.13 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, who were made inactive during the report quarter, or during the quarter after the report quarter (resolved, neither secured nor timely).

2.14 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, whose liability date (met threshold) was changed from prior to the report quarter, to during or after the report quarter (resolved, neither secured nor timely).

2.15 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, who were suspended from filing required reports due in the report quarter by virtue of being seasonal employers, an administrative decision not to pursue report filing, or for other reasons (resolved, neither secured nor timely).

2.16 Reimbursable employers owing required reports for activities in the quarter prior to the report quarter, whose accounts were withdrawn by making the liability date and the inactive/terminated "as of" date equal (resolved, neither secured nor timely). This includes canceled, withdrawn, closed, dropped, etc. accounts.

Table 3 Draft Reconstruction File Specifications
Status Determinations Entered within Report Quarter (RQ)

Report Quarter:
Validator:
Date:

These subpopulations constitute the unique subgroups of all status determinations made by the State during the ETA 581 Report Quarter (RQ). States that overwrite status determinations on their master tax file may use the RQC universe for reconstruction. Programmers and validators should note that time lapse categories are discrete subpopulations, whereas the ETA 581 reports time lapse cumulatively.

				1st Sort Key	2nd Sort Key									
Subpopulation	ETA 581 Item #'s	**1** *(Step 1C)* EAN	**2** *(Step 2A)* *(Step 2B)* Employer Type C/R	**3** *(Step 11A)* *(Step 11B)* *(Step 11C)* *(Step 11D)* Status Determ Type Indicator	**4** *(Step 12)* Time Lapse	**5** *(Step 13)* Status Determin. Date(s)	**6** *(Step 14)* Liability Date (Met Threshold)	**7** *(Step 14)* End of Liable Quarter	**8** *(Step 15)* Activation process date	**9** *(Step 16)* Reactivation process date(s)	**10** *(Step 17)* Successorship process date(s)	**11** *(Step 18)* Predecessor account number	**12** *(Step 6A)* *or* *(Step 6B)* Inactivation process date(s)	**13** *(Step 6A)* *or* *(Step 6C)* Termination process date(s)
3.1	14,15, 16			New	<=90 days	within RQ			within RQ, or <column 9 date	within RQ, or none			< active/ reactivation date, or blank	< active/ reactivation date, or blank
3.2	14,16			New	>=91 but <=180 days	within RQ			within RQ, or <column 9 date	within RQ, or none			< active/ reactivation date, or blank	< active/ reactivation date, or blank
3.3	14			New	>=181 days	within RQ			within RQ, or <column 9 date	within RQ, or none			< active/ reactivation date, or blank	< active/ reactivation date, or blank
3.4	17,18, 19			Successor	<=90 days	within RQ			<= successorship date	<= successorship date, or none	within RQ	non-blank		

Table 3 Draft Reconstruction File Specifications
Status Determinations Entered within Report Quarter (RQ)

				1st Sort Key	2nd Sort Key									
Subpopula-tion	ETA 581 Item #'s	**1** *(Step 1C)* EAN	**2** *(Step 2A)* *(Step 2B)* Employer Type C/R	**3** *(Step 11A)* *(Step 11B)* *(Step 11C)* *(Step 11D)* Status Determ Type Indicator	**4** *(Step 12)* Time Lapse	**5** *(Step 13)* Status Determin. Date(s)	**6** *(Step 14)* Liability Date (Met Threshold)	**7** *(Step 14)* End of Liable Quarter	**8** *(Step 15)* Activation process date	**9** *(Step 16)* Reactiva-tion process date(s)	**10** *(Step 17)* Succes-sorship process date(s)	**11** *(Step 18)* Prede-cessor account number	**12** *(Step 6A)* *or* *(Step 6B)* Inactiv-ation process date(s)	**13** *(Step 6A)* *or* *(Step 6C)* Termin-ation process date(s)
3.5	17,19			Successor	>=91 but <=180 days	within RQ			<= successor -ship date	<= successor -ship date, or none	within RQ	non-blank		
3.6	17			Successor	>=181 days	within RQ			<= successor -ship date	<= successor -ship date, or none	within RQ	non-blank		
3.7	20			Inactiv-ations	n/a	within RQ	n/a	n/a	n/a	n/a	n/a		within RQ[a]	blank
3.8	20			Termni-ations	n/a	within RQ	n/a	n/a	n/a	n/a	n/a		blank	within RQ

[a] There is the same issue as under Population #1, where the employer could be inactive based on 8 quarters of no wages (or fewer depending on the State's threshold), but for some reason the inactivation date/flag was not triggered. We may be able to cross-reference by EAN (by programming or on the printout) the employers identified as falling in this category from the Population #1 specifications, since they are identical, as long as the same RQ is validated.

Notes:

1) States that prefer to validate contributory and reimbursing employer status determinations separately may do so by replicating the eight subpopulations (one set of eight subpopulations for each type of employer). States may prefer to validate the two types of employers separately if they are processed in very different ways.

2. Time Lapse is the difference, in days, between the last day of the liable quarter and the status determination date.

Table 3 Draft Reconstruction File Specifications
Status Determinations Entered within Report Quarter (RQ)

Subpopulation Descriptions:

3.1 Status determinations of new employers made during the report quarter, which were made within 90 days of the end of the quarter in which the employer became liable. (Employers changing from contributory to reimbursing status and vice versa are included in subpopulations 3.1 - 3.3.)

3.2 Status determinations of new employers made during the report quarter, which were made between 91 and 180 days of the end of the quarter in which the employer became liable.

3.3 Status determinations of new employers made during the report quarter, which were made 181 days or later from the end of the quarter in which the employer became liable.

3.4 Status determinations of successor employers made during the report quarter, which were made within 90 days of the end of the quarter in which the employer became liable.

3.5 Status determinations of successor employers made during the report quarter, which were made between 91 and 180 days of the end of the quarter in which the employer became liable.

3.6 Status determinations of successor employers made during the report quarter, which were made 181 days or later from the end of the quarter in which the employer became liable.

3.7 Inactivations of employers made during the report quarter.

3.8 Terminations of employers made during the report quarter.

Table 4 Draft Reconstruction File Specifications
Accounts Receivable

Report Quarter:

Validator:

Date:

Sub-population	1 *(Step 1D)* EAN	2 *(Step 2A)* *(Step 2B)* Employer Type C/R	3 *(Step 19A)* Transaction Date	4 *(Step 19B)* Established Q/Date	5 *(Step 1D)* Employer Report Quarter (ERQ)	6 *(Step 20)* Due Date (DD)	7 *(Step 21A)* *(Step 21B)* *(Step 21C)* Transaction Type/Indicator	8 *(Step 22)* Amount Established in RQ	9 *(Step 23)* Liquidated (Pay/Adj)	10 *(Step 24)* Uncollectible	11 *(Step 25)* Removed	12 *(Step 26)* Balance at end of RQ	13 *(Step 27A)* *(Step 27B)* Age
4.1		C		RQ		blank	Establishment	$	blank	blank	blank	blank	blank
4.2		C	RQ	blank			Liquidation	blank	$	blank	blank	blank	blank
4.3		C	RQ	blank	>RQ-8	blank	Uncollectible	blank	blank	$	blank	blank	blank
4.4		C	RQ	>RQ-3	<RQ-8	blank	Uncollectible	blank	blank	$	blank	blank	blank
4.5		C	blank	RQ-2	RQ-8	blank	blank	blank	blank	blank	$	blank	blank
4.6		C	blank	RQ-3	<RQ-8	blank	blank	blank	blank	blank	$	blank	blank
4.7		C	blank	blank	>RQ-8	blank	blank	blank	blank	blank	blank	$	
4.8		C	blank	>RQ-3	<RQ-8	blank	blank	blank	blank	blank	blank	$	
ETA Item #								22	23	24	25	26	

Table 4 Draft Reconstruction File Specifications
Accounts Receivable

Sub-population	**1** *(Step 1D)* EAN	**2** *(Step 2A)* *(Step 2B)* Employer Type C/R	**3** *(Step 19A)* Transaction Date	**4** *(Step 19B)* Established Q/Date	**5** *(Step 1D)* Employer Report Quarter (ERQ)	**6** *(Step 20)* Due Date (DD)	**7** *(Step 21A)* *(Step 21B)* *(Step 21C)* Transaction Type/Indicator	**8** *(Step 22)* Amount Established in RQ	**9** *(Step 23)* Liquidated (Pay/Adj)	**10** *(Step 24)* Uncollectible	**11** *(Step 25)* Removed	**12** *(Step 26)* Balance at end of RQ	**13** *(Step 27A)* *(Step 27B)* Age
4.9		R		RQ		blank	Establishment	$	blank	blank	blank	blank	blank
4.10		R	RQ	blank			Liquidation	blank	$	blank	blank	blank	blank
4.11		R	RQ	blank	blank	>RQ-7	Uncollectible	blank	blank	$	blank	blank	blank
4.12		R	RQ	>RQ-3	blank	<RQ-7	Uncollectible	blank	blank	$	blank	blank	blank
4.13		R	blank	RQ-2	blank	RQ-7	blank	blank	blank	blank	$	blank	blank
4.14		R	blank	RQ-3	blank	<RQ-7	blank	blank	blank	blank	$	blank	blank
4.15		R	blank	blank	blank	>RQ-7	blank	blank	blank	blank	blank	$	
4.16		R	blank	>RQ-3	blank	<RQ-7	blank	blank	blank	blank	blank	$	
ETA Item #								34	35	36	37	38	

Table 4 Draft Reconstruction File Specifications
Accounts Receivable

Notes:

Values in column 8 for all observations in subpopulations 4.1 - 4.8 should be totaled, for comparison to ETA Item #22.
Values in column 9 for all observations in subpopulations 4.1 - 4.8 should be totaled, for comparison to ETA Item #23.
Values in column 10 for all observations in subpopulations 4.1 - 4.8 should be totaled, for comparison to ETA Item #24.
Values in column 11 for all observations in subpopulations 4.1 - 4.8 should be totaled, for comparison to ETA Item #25.
Values in column 12 for all observations in subpopulations 4.1 - 4.8 should be totaled, for comparison to ETA Item #26.

Values in column 8 for all observations in subpopulations 4.9 - 4.16 should be totaled, for comparison to ETA Item #34.
Values in column 9 for all observations in subpopulations 4.9 - 4.16 should be totaled, for comparison to ETA Item #35.
Values in column 10 for all observations in subpopulations 4.9 - 4.16 should be totaled, for comparison to ETA Item #36.
Values in column 11 for all observations in subpopulations 4.9 - 4.16 should be totaled, for comparison to ETA Item #37.
Values in column 12 for all observations in subpopulations 4.9 - 4.16 should be totaled, for comparison to ETA Item #38.

Subpopulation descriptions:

4.1 Receivable amounts established as past due in the report quarter for contributory employers

4.2 Receivable amounts liquidated during the report quarter for contributory employers

4.3 Receivable amounts declared uncollectible during the report quarter for contributory employers where the receivable is less than eight quarters old

Table 4 Draft Reconstruction File Specifications
Accounts Receivable

4.4 Receivable amounts declared uncollectible during the report quarter for contributory employers where the receivable is more than seven quarters old but was established within the report quarter or the two preceding quarters

4.5 Receivables removed during the report quarter for contributory employers where the receivable is eight quarters old and was established prior to two quarters before the report quarter

4.6 Receivables removed during the report quarter for contributory employers where the receivable was more than eight quarters old and was established three quarters prior to the report quarter

4.7 Receivable balances at the end of the report quarter for contributory employers which were less than eight quarters old

4.8 Receivable balances at the end of the report quarter for contributory employers which were more than eight quarters old but which were established within the report quarter or the two preceding quarters

4.9 Receivable amounts established as past due in the report quarter for reimbursable employers

4.10 Receivable amounts liquidated during the report quarter for reimbursable employers

4.11 Receivable amounts declared uncollectible during the report quarter for reimbursable employers where the receivable is less than eight quarters old

4.12 Receivable amounts declared uncollectible during the report quarter for reimbursable employers where the receivable is more than seven quarters old but was established within the report quarter or the two preceding quarters

Table 4 Draft Reconstruction File Specifications
Accounts Receivable

4.13 Receivables removed during the report quarter for reimbursable employers where the receivable is eight quarters old and was established prior to two quarters before the report quarter

4.14 Receivables removed during the report quarter for reimbursable employers where the receivable was more than eight quarters old and was established three quarters prior to the report quarter

4.15 Receivable balances at the end of the report quarter for reimbursable employers which were less than eight quarters old

4.16 Receivable balances at the end of the report quarter for reimbursable employers which were more than eight quarters old but which were established within the report quarter or the two preceding quarters.

Table 5 Draft Reconstruction File Specifications
Field audits

Report Quarter:

Validator:

Date:

These subpopulations constitute the unique subgroups of all field audits completed during the ETA 581 Report Quarter (RQ). Data elements specified on the file specification may not be captured on the State's system when they are not reported on the 581. They are however included in the auditor's file. When States cannot capture such information automatically, the column can be completed from the auditor's paper files during the validation for the selected cases.

				1st Sort Key	2nd Sort Key	3rd Sort Key	Total Wages					Taxable Wages					Contributions				
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
		(Step 1E)	*(Step 1E)*	*(Step 28A)* *(Step 28B)*	*(Step 29A)* *(Step 29B)*	*(Step 30)*	*(Step 31A)*	*(Step 31B)*	*(Step 31C)*	*(Step 31D)*	*(Step 31E)*	*(Step 32A)*	*(Step 32B)*	*(Step 32C)*	*(Step 32D)*	*(Step 32E)*	*(Step 33A)*	*(Step 33B)*	*(Step 33C)*	*(Step 33D)*	*(Step 33E)*
Sub-population	ETA 581 Items #'s	EAN	audit ID #	employer size L/S	change audit Y/N	audit completion date	pre	post	under	over	total reconcil-iation amount	pre	post	under	over	tax reconcil-iation amount	pre	post	under	over	cont. reconcil-iation amount
5.1	45, 46, 47			L	Y	within RQ	T1$	T2$	T3$	T4$	0	X1$	X2$	X3$	X4$	0	C1$	C2$	C3$	C4$	0
5.2	45, 47			L	N	within RQ															
5.3	46, 547			S	Y	within RQ															
5.4	47			S	N	within RQ															
ETA Item #							49	50	53	56				54	57				55	58	

Notes:

1) Some States may want to capture and store in the data file the number of employees pre-and post-audit. Some States allocate a percentage of their UI receipts to special funds or programs; if so, the employer's discount rate and amount discounted should be included on the printout.

Table 5 Draft Reconstruction File Specifications
Field audits

2) Post audit figures for total wages, taxable wages and contributions reflect the net increase or decrease of under and over reporting identified during the audit, even though the netted figures are not reportable on the ETA 581. Referring to the specification:

Subtract the positive net of (T3 - T4) from the positive net of (T1 - T2). The result in column 10 should be zero.

Subtract the positive net of (X3 - X4) from the positive net of (X1 - X2). The result in column 15 should be zero.

Subtract the positive net of (C3 - C4) from the positive net of (C1 - C2). The result in column 20 should be zero.

For example, if Employer A under reported total wages by $5000 and also over reported total wages by $1000, the Employer's post-audit total wages would increase by $4000. So, if the validator nets the under and over reported wages the result is $4000, and nets pre and post audit wages the result is $4000. These two results should always reconcile to zero. Referring again to the printout specification:

If TI = $10,000, T2 = $14,000, T3 = $5,000, T4 = $1,000, then ($10,000 - $14,000) - ($5,000 - $1,000) = 0

Also, if TI = $10,000, T2 = $6,000, T3 = $1,000, T4 = $5,000, then ($10,000 - $6,000) - ($1,000 - $5,000) = 0

3) The number of observations in all four subpopulations should be totaled, for comparison to ETA Item #47.

Dollar values in column 6 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #49.

Dollar values in column 7 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #50.

Dollar values in column 8 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #53.

Dollar values in column 9 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #56.

Dollar values in column 13 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #54.

Dollar values in column 14 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #57.

Dollar values in column 18 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #55.

Dollar values in column 19 for all observations in all four subpopulations should be totaled, for comparison to ETA Item #58.

Table 5 Draft Reconstruction File Specifications
Field audits

Subpopulation descriptions:

5.1 Large employer audits completed during the report quarter, which were change audits.

5.2 Large employer audits completed during the report quarter, which were not change audits.

5.3 Small employer audits completed during the report quarter, which were change audits.

5.4 Small employer audits completed during the report quarter, which were not change audits.

APPENDIX A

PART II

SORT SPECIFICATIONS

Part II of Appendix A contains sort specifications for report validation (RV) files and the criteria for review of the files once sorted. The reason the State sorts the files is to quickly identify transactions that have inappropriately been included in the RV files, transactions that are outside the parameters for inclusion in the population or subpopulation. For example, many transactions are supposed to be counted on the ETA 581 report only if they occur during a particular quarter. If the transactions are sorted in ascending order by the occurrence date the earliest dates will appear at the start of the list; the latest dates at the end. By looking at the first and last transactions in the sorted file the validators can quickly determine if all transactions are within range, occurring in the report quarter.

- The first column of the tables in Appendix A, Part II identifies the sort. The sorts with an "S" following the first digit are the sort specifications already identified by the "Sort Keys" in Part I of Appendix A. If the RV file was sorted according to the specifications there is no additional sorting required for these File Sorts. The sorts with a letter other than S following the first digit do require sorting the RV file in a different order.

- The second column identifies the subpopulations to be sorted. Some sorts include transactions for more than one subpopulation.

- The Sort Key indicates the data element(s) to be used when sorting the file into the new order.

- The Error Detection Criteria provide instructions on how to review the sorted file. The review will tell the State staff whether all transactions in the file are within the specified range. The Step number refers the State staff to a location in the State-specific information in Module 3 of the handbook where they can find additional clarification of the information used in the sort.

Population 1 Sort Criteria - Active Employers

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
1.S.1	1.1 through 1.2	Employer Status Indicator (all observations combined)	3A/1	All employer status indicators must represent active employers. (If the State has several codes for active, the validator must look at every break between codes to ensure that all employer status indicators represent active.) This ensures that inactive employers are not included in these subpopulations. The indicator cannot be potentially subject, inactive or suspended.
1.S.2	1.1 through 1.2	Employer Type (all observations within each subpopulation)	2A/1	All type codes in subpopulation 1.1 must represent contributory employers. This ensures that reimbursing employers are not included in these subpopulations.
1.S.3			2B/1	All type codes in subpopulation 1.2 must represent reimbursing employers. This ensures that contributory employers are not included in these subpopulations.

Population 1 *(continued)*

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
1.S.4[1]	1.1 through 1.2	Liability Date (reopen) (all observations with a liability date (reopen) within each subpopulation in descending order)	3A/2	All liability dates (reopen) must be prior to or within the report quarter. This ensures that employers who are potentially subject or inactive are not included in these subpopulations.
		Liability Date (initial) (all remaining observations in each subpopulation)		All liability dates (initial) must be prior to or within the report quarter. This ensures that employers who are potentially subject are not included in these subpopulations.

[1] Sort 1.S.4 is a two-tiered sort. The first sort key is liability date (reopen) and the second sort key, on transactions within each subpopulation that do not have a liability date (reopen), is liability date (initial).

NOTE: Sorts for each population begin with the number of the corresponding population table. Sorts where the second character is the letter "S" are included on the reconstruction file sort. The Step No./Rule No. column refers to the Data Element Validation Instruction in Module 3 which corresponds to the sort.

Population 1 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
1.A	1.1 (active contributory employers)	EAN (ascending by prefix or suffix)	2A/2	All EAN prefixes or suffixes (whichever method a State uses, if either, to distinguish employer types) must represent contributory employers. This ensures that reimbursing employers are not included in this subpopulation.
1.B	1.2 (active reimbursing employers)	EAN (ascending by prefix or suffix)	2B/2	All EAN prefixes or suffixes (whichever method a State uses, if either, to distinguish employer types) must represent reimbursing employers. This ensures that contributory employers are not included in this subpopulation.
1.C	1.1 through 1.2 (all active employers)	Inactive/Terminated "as of" Date (Sort non-blank dates only in ascending order)	3A/3	All inactive/terminated "as of" dates, where they exist, must be after the report quarter. Draw a sample of 200 inactive/terminated "as of" dates which are not after the report quarter; the validator reviews these to ensure that the inactive/terminated "as of" date is before the liability date (reopen). This ensures that employers inactive on the last day of the report quarter are not included in these subpopulations.

Population 1 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
1.D[1]	1.1 through 1.2 (all active employers)	Number of Quarters (Primary sort key, descending)	3A/4	The sum of wages must be greater than zero if the number of quarters is eight.
		Sum of Wages Reported (Secondary sort key, ascending)		

[1] Sort 1.D is a two-tiered sort. Observations in the specified subpopulations are first sorted in descending order by the Number of Quarters. Then, within each grouping by number of quarters (e.g. all with eight quarters, all with seven quarters), observations are sorted in ascending order by the Sum of Wages Reported. This approach will group all of the observations with the Number of Quarters at eight and a Sum of Wages of zero together at the top. These are the invalid observations.

Population 2 Sort Criteria - Report Filing

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
2.S.1	2.1 through 2.16 (all employer reports)	Employer Type	2A/1	All type codes in subpopulations 2.1 through 2.8 must represent contributory employers. This ensures that reimbursing employers are not included in these subpopulations.
2.S.2			2B/1	All type codes in subpopulations 2.9 through 2.16 must represent reimbursing employers. This ensures that contributory employers are not included in these subpopulations.
2.S.3	2.1 through 2.16 (all employer reports)	Filing Status Indicator (descending alphabetically if T,S,R) (Sort may be n/a if indicator is not a true value.)	8A/1	All filing status indicators in subpopulations 2.1 and 2.9 must represent timely filed reports, to ensure that untimely reports are not included in these subpopulations.
2.S.4			8B/1	All filing status indicators in subpopulations 2.2 and 2.10 must represent only secured reports. This ensures that no timely filed or resolved reports are included in these subpopulations.
2.S.5			8C/1	All filing status indicators in subpopulations 2.3 through 2.8 and 2.11 through 2.16 must represent only resolved reports. This ensures that no timely or secured reports are included in these subpopulations.

Population 2 *(continued)*

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
2.S.6	2.1 through 2.16 (all employer reports)	Received Date (ascending when non-blank)	8A/2	All received dates in subpopulations 2.1 and 2.9 must be prior to or equal to the due date. This ensures that untimely filed reports are not included in these subpopulations.
2.S.7			8B/2	All received dates in subpopulations 2.2 and 2.10 must be after the due date and by the end of the report quarter. This ensures that timely or resolved (neither timely nor secured) filed reports are not included in these subpopulations.
2.S.8			8C/2	All received dates in subpopulations 2.3 and 2.11 must be during the quarter following the report quarter. This ensures that timely or secured filed reports are not included in these subpopulations.

Population 2 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
2.A	2.1 through 2.8	EAN (ascending by prefix or suffix)	2A/2	All EAN prefixes or suffixes (whichever method a State uses, if either, to distinguish employer types) for subpopulations 2.1 through 2.8 must represent contributory employers. This ensures that reimbursing employers are not included in these subpopulations.
2.B	2.9 through 2.16	EAN (ascending by prefix or suffix)	2B/2	All EAN prefixes or suffixes (whichever method a State uses, if either, to distinguish employer types) for subpopulations 2.9 through 2.16 must represent reimbursing employers. This ensures that contributory employers are not included in these subpopulations.
2.C	2.4, 2.12 (reports resolved by final assessment)	Final Assessment Date (descending)	8C/3	All final assessment dates must be by the end of the quarter following the report quarter. This ensures that reports resolved later by final assessment are not included in these subpopulations.
2.D	2.6, 2.14	Liability Date (met threshold) (ascending)	8C/4	All liability dates must be during or after the report quarter. This ensures that only employers whose reports were resolved by changing the liability date (met threshold) are included in these subpopulations.
2.E	2.8, 2.16	Liability Date (initial or reopen) (ascending)	8C/5	All liability dates (initial or reopen) must be equal to the inactive/terminated "as of" date. Validator reviews all dates to ensure that they are equal. This ensures that only employers whose reports were resolved by canceling, withdrawing, closing, or dropping the account are included in these subpopulations. (Some States, such as Colorado, may have a slight variation from this criteria.)

Population 2 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
2.F	2.5, 2.13	Inactive/Terminated "as of" Date (descending)	8C/6	All inactive/terminated "as of" dates must be prior to the quarter before the report quarter, to ensure that only employers whose reports were resolved by inactivating the employer are included in these subpopulations.
2.G	2.7, 2.15	Suspended "as of" quarter (ascending)	8C/7	All suspended "as of" quarters, where they exist, must be equal to the quarter before the report quarter (the ERQ), to ensure that only employers who were suspended from filing an employer report during the report quarter are included in these subpopulations.
2.H	2.5, 2.13	Inactivation/ Termination Processing Date (ascending)	8C/8	All inactivation/termination processing dates must be after the first day of the quarter before the report quarter and by the last day of the quarter after the report quarter, to ensure that only employers whose reports were resolved by inactivating the employer are included in these subpopulations.

Population 3 Sort Criteria - Status Determinations

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
3.S.1	3.1 through 3.8 (all status determinations in report quarter)	Status Determination Type Indicator	11A/1	All Status Determination Type Indicators in subpopulations 3.1 through 3.3 must represent new employer determinations.
3.S.2			11B/1	All Status Determination Type Indicators in subpopulations 3.4 through 3.6 must represent successor employer determinations.
3.S.3			11C/1	All Status Determination Type Indicators in subpopulation 3.7 must represent inactivated employer determinations.
3.S.4			11D/1	All Status Determination Type Indicators in subpopulation 3.8 must represent terminated employer determinations.
3.S.5	3.1 through 3.6 (all time lapsed status determinations)	Time Lapse (ascending)	12/1	All time lapse counts in subpopulations 3.1 and 3.4 must be less than or equal to 90 days. All time lapse counts in subpopulations 3.2 and 3.5 must be greater than or equal to 91 days, but less than or equal to 180 days. All time lapse counts in subpopulations 3.3 and 3.6 must be greater than or equal to 181 days. Validator should also examine time lapse outliers as follows: first five transactions in subpopulations 3.1 and 3.4, last five transactions in 3.3 and 3.6.

Population 3 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
3.A	3.1 through 3.8 (all status determinations in report quarter)	Status Determination Date (ascending)	13/1	All status determination dates must be within the report quarter being validated, to ensure that all status determinations included in these subpopulations are reportable in the same ETA 581 report quarter. (States may not have the capacity to perform this sort if their systems overwrite each status determination attributable to an employer.) If State maintains specific processing dates and uses/copies those dates here, this can replace sorts 3.B - 3.D, 3.F, 3.G.
3.B[1]	3.1 through 3.3 (all new status determinations)	Activation Processing Date (ascending)	11A/2	All activation processing dates must be during the report quarter, unless an observation has a reactivation processing date during the report quarter.
		Reactivation Processing Date (ascending)		Resort transactions with activation processing dates prior to the report quarter by reactivation processing date. All reactivation processing dates, where they exist, must be during the report quarter.
3.C	3.4 through 3.6 (all successor status determinations)	Successorship Processing Date (ascending)	11B/3	All successorship processing dates must be during the report quarter.
3.D	3.4 through 3.6 (all successor status determinations)	Predecessor Account Number (ascending)	11B/2	All observations must have a predecessor account number, to ensure that all observations represent successors. (Only applicable if State requires maintenance of predecessor numbers.)

[1] Sort 3.B is a two-tiered sort, with the primary sort key activation processing date.

Population 3 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
3.E	3.7 (all inactivation status determinations)	Inactivation Processing Date (ascending)	11C/2	All inactivation processing dates must be during the report quarter.
3.F	3.8 (all termination status determinations)	Termination Processing Date (ascending)	11D/2	All termination processing dates must be during the report quarter.

States that use EAN prefix or suffix to identify status determination type may also sort by EAN by subpopulation groupings 3.1 - 3.3, 3.4 - 3.6, and 3.7 - 3.8.

Population 4 Sort Criteria - Accounts Receivable

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
4.A	4.2 and 4.10	Transaction Date	23/1	All transaction dates must be during the report quarter.
	4.3, 4.4, 4.11 and 4.12		24/1	All transaction dates must be during the report quarter.
4.B	4.1 and 4.9	Established Quarter/Date	22/1	All established dates must be during the report quarter.
4.C	4.1 and 4.9	Transaction Type/Indicator	22/2	All transaction type indicators represent amounts established.
4.D	4.1 and 4.9	Amount Established in Report Quarter	22/3	All established amounts are greater than zero.
4.E	4.2 and 4.10	Transaction Type/Indicator	23/2	All transaction type indicators represent amounts liquidated.
4.F	4.2 and 4.10	Liquidated (Pay/Adj)	23/3	All liquidated amounts are greater than zero.
4.G	4.4 and 4.12	Established Quarter/Date	24/2	All established dates must be during the report quarter or the two preceding quarters.
	4.5 and 4.13		25/1	All established dates must be during the report quarter or the two preceding quarters.
4.H	4.4	Employer Report Quarter	24/3	All employer report quarters must be at least eight quarters prior to the validation report quarter.
	4.6		25/4	All employer report quarters must be at least eight quarters prior to the validation report quarter.
	4.8		26/3	All employer report quarters must be at least eight quarters prior to the validation report quarter.

Population 4 *(continued)*

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
4.I	4.3	Employer Report Quarter	24/4	All employer report quarters must be less than eight quarters prior to the validation report quarter.
	4.7		26/2	All employer report quarters must be less than eight quarters prior to the validation report quarter.
4.J	4.12	Due Date	24/5	All employer report due dates must be more than seven quarters prior to the report quarter.
	4.14		25/6	All employer report due dates must be more than seven quarters prior to the report quarter.
	4.16		26/5	All employer report due dates must be more than seven quarters prior to the report quarter.
4.K	4.11	Due Date	24/6	All employer report due dates must be less than eight quarters prior to the validation report quarter.
	4.15		26/4	All employer report due dates must be less than eight quarters prior to the validation report quarter.
4.L	4.3, 4.4, 4.11, and 4.12	Transaction Type/Indicator	24/7	All transaction type indicators represent amounts declared uncollectible.
4.M	4.3, 4.4, 4.11, and 4.12	Uncollectible	24/8	All uncollectible amounts are greater than zero.
4.N	4.6 and 4.14	Established Quarter/Date	25/2	All established dates are three quarters prior to the validation report quarter.
4.O	4.5	Employer Report Quarter	25/3	All employer report quarters are eight quarters prior to the validation report quarter.

Population 4 *(continued)*

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
4.P	4.13	Due Date	25/5	All employer report due dates are seven quarters prior to the validation report quarter.
4.Q	4.5, 4.6, 4.13, and 4.14	Removed	25/7	All removed amounts are greater than zero.
4.R	4.8	Established Quarter/Date	26/1	All established dates are within the report quarter or the two preceding report quarters.
4.S	4.7, 4.8, 4.15, and 4.16	Balance at end of Report Quarter	26/6	All end of quarter balances arae greater than zero.

Population 5 Sort Criteria - Field Audits

Reconstruction File Sort	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
5.S.1	5.1 through 5.4 (all audits completed in report quarter)	Employer Size	28A/1	All indicators in subpopulations 5.1 and 5.2 must represent large employers, to ensure that small employers are not included in these subpopulations. .
5.S.2			28B/1	All indicators in subpopulations 5.3 and 5.4 must represent small employers, to ensure that large employers are not included in these subpopulations
5.S.3	5.1 through 5.4 (all audits completed in report quarter)	Change Audit	29A/1	All indicators in subpopulations 5.1 and 5.3 must represent “change” audits, to ensure that “no change” audits are not included in these subpopulations.
5.S.4			29B/1	All indicators in subpopulations 5.2 and 5.4 must represent “no change” audits, to ensure that “change” audits are not included in these subpopulations.
5.S.5	5.1 through 5.4 (all audits completed in report quarter)	Audit Completion Date	30/1	All audit sign-off dates should be during the report quarter, to ensure that all audits included in these subpopulations are reportable in the same ETA 581 report quarter.

Population 5 *(continued)*

Sort Number	Subpopulations Included in Sort	Sort Key/ Column No.	Step No./ Rule No.	Error Detection Criteria
5.A	5.1 through 5.4 (all audits completed in report quarter)	Total Wages Reconciliation (descending)	31E/1	All total wages reconciliation amounts should equal zero.
5.B	5.1 through 5.4 (all audits completed in report quarter)	Taxable Wages Reconciliation (descending)	32E/1	All taxable wages reconciliation amounts should equal zero.
5.C	5.1 through 5.4 (all audits completed in report quarter)	Contributions Reconciliation (descending)	33E/1	All contributions reconciliation amounts should equal zero.

Sorts 5.B and 5.C rely on data which the States may not capture on their systems, and which may have to be obtained from paper files. It is intended that this manually stored data would be captured at the time of validation for sampled transactions only. These sorts require that the data be captured for all observations much earlier in the process and therefore can only be performed by States with that system capability.

•Appendix B

REPORT ITEM VALIDATION SPREADSHEETS

INTRODUCTION

Appendix B contains facsimiles of the spreadsheets States will use to enter and report their validation findings. States will receive an electronic version of their spreadsheets. After States enter their findings, the spreadsheets automatically calculate the difference between the validation counts and the information reported on the ETA 581 report.

Data needs to be entered only in a few cells.

1. If, under the column labeled "ETA 581 Item", there is a number identifying a cell on the ETA 581 report enter the information reported in that cell. The information should be entered on the corresponding row under the column labeled "Reported Count".

2. If there is a number under the "Subpopulation" column enter the validation count for that subpopulation in the column with the heading "Validation Count".

3. If data validation procedures call for a review of duplicates there will be a cell at the bottom of the spreadsheet titled "Duplicate Calculations". Enter the number of transactions reviewed and the number of duplicates. The spreadsheet will calculate the percent of transactions that are duplicates in the sample. It will adjust overall validation counts for these findings.

4. For each sort enter the "Sort Count" (typically the total number of transactions in the sorted subpopulations) and the "Sort Count Difference" the number of cases that were out of range.

5. There is a space where states can enter comments regarding the causes for errors found and any action they plan in response to the errors.

REPORT VALIDATION SUMMARY SPREADSHEET: POPULATION #1

	ETA 581 ITEM	ETA REPORTED COUNT	SUBPOPULATION	SUBPOPULATION VALIDATION COUNT	% DUPL. ERRORS	ADJUSTED VALIDATION COUNT	COUNT DIFFERENCE	COUNT % DIFFERENCE	COUNT PASS/ FAIL
Active contributory employers	1		1.1				0		
Active reimbursing employers	2		1.2				0		
All active employers	3		TOTAL	**0**		**0.00**	0		**PASS**

Duplicate Calculations	
Transactions Reviewed	# of Duplicates

SORT NUMBER	SORT COUNT	SORT COUNT DIFFERENCE
1.S.1		0
1.S.2		0
1.S.3		0
1.S.4		0
1.A		0
1.B		0
1.C		0
1.D		0

NOTES: COUNT % DIFFERENCE = (REPORTED COUNT (RC) - VALIDATION COUNT (VC))/VC for the TOTAL row.. If the COUNT % Difference in any row is greater than +/- 2%,
the COUNT PASS/FAIL column will indicate a FAIL. In the applicable populations, the $ % DIFFERENCE and $ PASS/FAIL are calculated in the same way that COUNT figures are calculated.
If duplicate counts are found in populations in which dollar figures are validated, the VALIDATION $ Amount will be inflated as a result. It is not possible to adjust VALIDATION $ on the basis of duplicate counts.

SUBPOPULATION		COMMENTS
1.1	Active contributory employers	
1.2	Active reimbursing employers	
Total	All active employers	

REPORT VALIDATION SUMMARY SPREADSHEET: POPULATION #2

Employers Owing Reports for Prior Quarter	ETA 581 ITEM	ETA REPORTED COUNT	SUBPOPULATION	SUBPOPULATION VALIDATION COUNT	% DUPL. ERRORS	ADJUSTED VALIDATION COUNT	COUNT DIFFERENCE	COUNT % DIFFERENCE	COUNT PASS/ FAIL	SORT NUMBER	SORT COUNT	SORT COUNT DIFFERENCE
Contributory Employers										2.S.1		0
Timely	6		2.1				0	#DIV/0!		2.S.2		0
Secured, not timely	Item 7 - Item 6	0	2.2				0	#DIV/0!		2.S.3		0
Total Secured	7		Sum of 2.1 + 2.2	0			0	#DIV/0!		2.S.4		0
Resolved - filed			2.3							2.S.5		0
Resolved - assessment			2.4							2.S.6		0
Resolved - inactive			2.5							2.S.7		0
Resolved - new liab. date			2.6							2.S.8		0
Resolved - suspended			2.7							2.A		0
Resolved - Cancel			2.8							2.B		0
Resolved, not secured, not timely	Items(8 in RQ+1)-7-6	0	Sum of 2.3 - 2.8	0			0			2.C		0
Total Resolved	Item 8 in RQ+1		Sum of 2.1 - 2.8	0			0			2.D		0
Reimbursable Employers										2.E		0
Timely	9		2.9				0			2.F		0
Secured, not timely	Item 10- Item 9	0	2.10				0			2.G		0
Total Secured	10		Sum of 2.9 +2.10	0			0			2.H		0
Resolved - filed			2.11									
Resolved - assessment			2.12									
Resolved - inactive			2.13									
Resolved - new liab. date			2.14									
Resolved - suspended			2.15									
Resolved - Cancel			2.16									
Resolved, not secured, not timely	Item 11(in RQ+1)-10-9	0	Sum of 2.11 - 2.16	0			0					
Total Resolved	11 in RQ+1		Sum of 2.9 - 2.16	0			0					
			TOTAL	0		0.00	0		**PASS**			

Duplicate Calculations	
Transactions Reviewed	# of Duplicates

NOTES: COUNT % DIFFERENCE = (REPORTED COUNT (RC) - VALIDATION COUNT (VC))/VC. If the COUNT % Difference in any row is greater than +/- 2%, the COUNT PASS/FAIL column will indicate a FAIL. In the applicable populations, the $ % DIFFERENCE and $ PASS/FAIL are calculated in the same way that COUNT figures are calculated.
If duplicate counts are found in populations in which dollar figures are validated, the VALIDATION $ Amount will be inflated as a result. It is not possible to adjust VALIDATION $ on the basis of duplicate counts.

SUBPOPULATION		COMMENTS
Contributory Employers		
2.1	Timely	
2.2	Secured, not timely	
Sum of 2.1 + 2.2	Total Secured	
2.3	Resolved - assessment	
2.4	Resolved - inactive	
2.5	Resolved - new liab. date	
2.6	Resolved - suspended	
2.7	Resolved - Cancel	
2.8	Resolved, not secured, not timely	
Sum of 2.3 - 2.8	Total Resolved	
Sum of 2.1 - 2.8	Reimbursable Employers	
Reimbursable Employers		
2.9	Timely	
2.10	Secured, not timely	
Sum of 2.9 +2.10	Total Secured	
2.1	Resolved - filed	
2.1	Resolved - assessment	
2.1	Resolved - inactive	
2.1	Resolved - new liab. date	
2.2	Resolved - suspended	
2.2	Resolved - Cancel	
Sum of 2.11 - 2.16	Resolved, not secured, not timely	
Sum of 2.9 - 2.16	Total Resolved	
TOTAL		

REPORT VALIDATION SUMMARY SPREADSHEET: POPULATION #3

STATUS DETERMINATIONS	ETA 581 ITEM	ETA REPORTED COUNT	SUBPOPULATION	SUBPOPULATION VALIDATION COUNT	% DUPL. ERRORS	ADJUSTED VALIDATION COUNT	COUNT DIFFERENCE	COUNT % DIFFERENCE	COUNT PASS/ FAIL	SORT NUMBER	SORT COUNT	SORT COUNT DIFFERENCE
Total New	14		Sum of 3.1+3.2+3.3	0			0					
New, <= 90 days	15		3.1				0			3.S.1		0
New, 91-180 days			3.2							3.S.2		0
New, <= 180 days	16		Sum of 3.1+ 3.2	0			0			3.S.3		0
New, 181 days or later			3.3				0			3.S.4		0
Total Successor	17		Sum of 3.4+3.5+3.6	0			0			3.S.5		0
Successor, <= 90 days	18		3.4				0			3A		0
Successor, 91-181 days			3.5							3B		0
Successor <= 180 days	19		Sum of 3.4+3.5	0			0			3C		0
Successor, 181 days or later			3.6							3D		0
Inactivations during RQ			3.7							3E		0
Terminations during RQ			3.8							3F		0
Inactive Terminations	20		Sum of 3.7 + 3.8	0			0					
			TOTAL	0		0.00	0		**PASS**			

Duplicate Calculations	
Transactions Reviewed	# of Duplicates

NOTES: COUNT % DIFFERENCE = (REPORTED COUNT (RC) - VALIDATION COUNT (VC))/VC. If the COUNT % Difference in any row is greater than +/- 2%,

the COUNT PASS/FAIL column will indicate a FAIL. In the applicable populations, the $ % DIFFERENCE and $ PASS/FAIL are calculated in the same way that COUNT figures are calculated.

If duplicate counts are found in populations in which dollar figures are validated, the VALIDATION $ Amount will be inflated as a result. It is not possible to adjust VALIDATION $ on the basis of duplicate counts.

SUBPOPULATION		COMMENTS
Sum of 3.1+3.2+3.3	Total New	
3.1	New, <= 90 days	
3.2	New, 91-180 days	
Sum of 3.1+ 3.2	New, <= 180 days	
3.3	New, 181 days or later	
Sum of 3.4+3.5+3.6	Total Successor	
3.4	Successor, <= 90 days	
3.5	Successor, 91-181 days	
Sum of 3.4+3.5	Successor <= 180 days	
3.6	Successor, 181 days or later	
3.7	Inactivations during RQ	
3.8	Terminations during RQ	
Sum of 3.7 + 3.8	Inactive Terminations	
TOTAL		

REPORT VALIDATION SUMMARY SPREADSHEET: POPULATION #4

	ETA 581 ITEM	ETA REPORTED COUNT	SUBPOPULATION	SUBPOPULATION VALIDATION COUNT	AMOUNT DIFFERENCE	% AMOUNT DIFFERENCE	PASS/ FAIL	SORT NUMBER	SORT COUNT	SORT COUNT DIFFERENCE
Conributory employers										
Established in RQ	22		4.1 Column 8		0		**PASS**	4.A		0
Liquidated in RQ	23		4.2 Column 9		0		**PASS**	4.B		0
Uncollectable <8Qs old			4.3 Column 10					4.C		0
Uncollectable >8Qs old			4.4 Column 10					4.D		0
Total Uncollectable	24		Sum of 4.3 and 4.4	0	0		**PASS**	4.E		0
Removed 8 Qs old, est > 2 Qs ago			4.5 Column 11					4.F		0
Balance >8Qs old, est RQ-3			4.6 Column 11					4.G		0
Total Removed	25		Sum of 4.5 and 4.6	0	0		**PASS**	4.H		0
Balance <8Qs old			4.7 Column 12					4.I		0
Balance ≥8Qs old, est after RQ-3			4.8 Column 12					4.J		0
Total Balance EOQ	26		Sum of 4.7 and 4.8	0	0		**PASS**	4.K		0
Reimbursable Employers								4.L		0
Established in RQ	34		4.9 Column 8		0		**PASS**	4.M		0
Liquidated in RQ	35		4.10 Column 9		0		**PASS**	4.N		0
Uncollectable <8Qs old			4.11 Column 10					4.O		0
Uncollectable >8Qs old			4.12 Column 10					4.P		0
Total Uncollectable	36		Sum of 4.11 and 4.12	0	0		**PASS**	4.Q		0
Removed 8 Qs old, est > 2 Qs ago			4.13 Column 11					4.R		0
Balance >8Qs old, est RQ-3			4.14 Column 11					4.S		0
Total Removed	37		Sum of 4.13 and 4.14	0	0	0.00%	**PASS**			
Balance <8Qs old			4.15 Column 12							
Balance ≥8Qs old, est after RQ-3			4.16 Column 12							
Total Balance EOQ	38		Sum of 4.15 and 4.16	0	0		**PASS**			

NOTES: COUNT % DIFFERENCE = (REPORTED COUNT (RC) - VALIDATION COUNT (VC))/VC. If the COUNT % Difference in any row is greater than +/- 2%, the COUNT PASS/FAIL column will indicate a FAIL. In the applicable populations, the $ % DIFFERENCE and $ PASS/FAIL are calculated in the same way that COUNT figures are calculated.

If duplicate counts are found in populations in which dollar figures are validated, the VALIDATION $ Amount will be inflated as a result. It is not possible to adjust VALIDATION $ on the basis of duplicate counts.

ITEMS		COMMENTS
Conributory employers		
Established in RQ	4.1 Column 8	
Liquidated in RQ	4.2 Column 9	
Uncollectable <8Qs old	4.3 Column 10	
Uncollectable >8Qs old	4.4 Column 10	
Total Uncollectable	Sum of 4.3 and 4.4	
Removed 8 Qs old, est > 2 Qs ago	4.5 Column 11	
Balance >8Qs old, est RQ-3	4.6 Column 11	
Total Removed	Sum of 4.5 and 4.6	
Balance <8Qs old	4.7 Column 12	
Balance >8Qs old, est after RQ-3	4.8 Column 12	
Total Balance EOQ	Sum of 4.7 and 4.8	
Reimbursable Employers		
Established in RQ	4.9 Column 8	
Liquidated in RQ	4.10 Column 9	
Uncollectable <8Qs old	4.11 Column 10	
Uncollectable >8Qs old	4.12 Column 10	
Total Uncollectable	Sum of 4.11 and 4.12	
Removed 8 Qs old, est > 2 Qs ago	4.13 Column 11	
Balance >8Qs old, est RQ-3	4.14 Column 11	
Total Removed	Sum of 4.13 and 4.14	
Balance <8Qs old	4.15 Column 12	
Balance >8Qs old, est after RQ-3	4.16 Column 12	
Total Balance EOQ	Sum of 4.15 and 4.16	

DRAFT

REPORT VALIDATION SUMMARY SPREADSHEET: POPULATION #5

	ETA 581 ITEM	ETA REPORTED COUNT	SUBPOPULATION	SUBPOPULATION VALIDATION COUNT	% DUPL. ERRORS	ADJUSTED VALIDATION COUNT	COUNT DIFFERENCE	COUNT % DIFFERENCE	COUNT PASS/ FAIL
Large, Change			5.1						
Large, no change			5.2						
Total large employer audits	45		Sum of 5.1 and 5.2	0			0		
Small, Change			5.3						
Small, no change			5.4						
Total change audits	46		Sum of 5.1 and 5.3	0			0		
Total audits	47		Sum 5.1 - 5.4	0	#DIV/0!	#DIV/0!	0		**PASS**

SORT NUMBER	SORT COUNT	SORT COUNT DIFFERENCE
5.S.1		0
5.S.2		0
5.S.3		0
5.S.4		0
5.S.5		0
5.A		0
5.B		0
5.C		0

Duplicate Calculations	
Transactions Reviewed	# of Duplicates

	ETA 581 ITEM	REPORTED $	SUBPOPULATION	VALIDATION $	$ DIFFERENCE	$ % DIFFERENCE	$ PASS/ FAIL
Total wages pre-audit	49		6				**PASS**
Total wages post-audit	50		7				**PASS**
Total wages under-reported	53		8				**PASS**
Total wages over-reported	56		9				**PASS**
Tax wages pre-audit							**PASS**
Tax wages post-audit							**PASS**
Tax wages under-reported	54		13				**PASS**
Tax wages over-reported	57		14				**PASS**
Contributions pre-audit							**PASS**
Contributions post-audit							**PASS**
Contributions under-reported	55		18				**PASS**
Contributions over-reported	58		19				**PASS**

	COMMENTS
Large, Change	
Large, no change	
Total large employer audits	
Small, Change	
Small, no change	
Total change audits	
Total audits	

	COMMENTS
Total wages pre-audit	
Total wages post-audit	
Total wages under-reported	
Total wages over-reported	
Tax wages pre-audit	
Tax wages post-audit	
Tax wages under-reported	
Tax wages over-reported	
Contributions pre-audit	
Contributions post-audit	
Contributions under-reported	
Contributions over-reported	

Appendix C

INDEPENDENT COUNT

APPENDIX C IS ONLY APPLICABLE TO POPULATIONS FOR WHICH THE STATE HAS PRODUCED THE RIV FILE FROM THE SAME EXTRACT FILES USED TO PRODUCE THE ETA 581 REPORT.

A. PURPOSE

The validation exercises described in Modules 1.1 and 1.2, and those outlined in Module 2, address the validation of all UI contributions transactions that have been *included* in the ETA 581 report. However, it is also important to confirm that no transactions have been improperly or systematically *excluded* from the Federal report. Although this problem is a difficult one, it is important to ensure that funding, economic statistics, and performance outcomes are not biased by the systematic elimination of particular types of transactions.

This module is not applicable for States that produce the RIV file directly from the employer contributions database, because the RIV process itself constitutes an independent count through the process of reconstruction. When the RIV file is produced from the same file used to produce the ETA 581 report, it is necessary to conduct an independent count in order to identify any errors that may have occurred in the ETA 581 report since these errors will be duplicated in the reconstruction file.

It is also not possible to perform an independent count when the database does not contain all of the reported transactions. In these circumstances, the statistical file is the only source of data to reconstruct reported counts on the ETA 581 report. It is unlikely that any State will need to perform an independent count for 581 validation as explained in Exhibit C.2 (it is more relevant to validating Federal benefits reports). This procedure is included in this handbook to ensure that States are aware of the possible problems with using statistical files for both reporting and validation when database files could be used.

B. MATERIALS (ADP STAFF)

1. Independent Count Files

ADP staff create independent total counts of transactions from the main database for comparison with counts generated on the extract files used to create the ETA 581. In

general, the independent count is created opposite to the way the RIV file is created. The RIV file should be programmed from the bottom up, by selecting only the codes and criteria indicated on the file specification in Appendix A. However, the independent count should be programmed from the top down, by including all codes relevant to a population and then subtracting observations related to those not indicated on the file specification.

Exhibit C.1 indicates when independent count validation is required. There are six typical scenarios for how States produce the ETA 581 report and reconstruct counts for validation. The ETA 581 Report Source column indicates for each scenario the source files that States use to generate report counts. States may use different source files for different types of transactions. The Validation Reconstruction Source column indicates for each scenario the source files that States use to reconstruct lists of transactions for validation.

The Independent Count Required column of Exhibit C.1 indicates whether the State should conduct independent count validation for transaction types that match the report and validation scenario.

Exhibit C.2 describes independent count criteria for each population.

Appendix D contains a spreadsheet to record the results of an independent count if one proves necessary.

APPENDIX C

EXHIBIT C.1

ETA 581 REPORTING AND VALIDATION CONFIGURATIONS

Scenario	Transactions Overwritten on Database	ETA 581			Data Validation			Independent Count Required
		Program Type	Source	Timing	Program Type	Source	Timing	
1	No	Count	Database	Snapshot (for reporting period)	Detail Record Extract (DRE)	Database	Snapshot	No
2	No	Count	Stat file	Daily	DRE	Database	Snapshot	No
3	No	DRE	Database	Snapshot (for reporting period)	DRE	Database	Snapshot	Yes
4	No	DRE	Stat file	Daily	DRE	Stat file	Daily	Yes
5	Yes	DRE	Stat file	Daily	DRE	Stat file	Daily	NA
6	Yes	Count	Stat file	Daily	must create a daily extract	NA	NA	NA

APPENDIX C

EXHIBIT C.2

INDEPENDENT COUNT CRITERIA, BY POPULATION (USING QUERY CAPABILITY)

Population Description	Independent Count Criteria
1 Active Employers	States should not use statistical files to validate active employers because the count should be taken from the database as a snapshot at the end of the month. If States do not use this approach for reporting (if they instead derive the number from changes in status over the quarter), they must use it for validation (they cannot recreate the active employer population from the status changes). Therefore, there is no situation which would require an independent count.
2 Report Filing	States generally use data files containing a record for each employer quarter for both reporting and reconstructing counts of employer report statuses. Therefore, there is not likely to be a situation where statistical files are used for reporting or validation. If a State uses a statistical file for validation, it should create a frequency distribution of received dates for every employer with a received date for the quarter being validated. This count can be used to validate that the statistical file data matches the data base for all timely and secured reports and for all reports which are resolved by receipt of report. This will validate subpopulations 2.1, 2.2, 2.3, 2.9, 2.10 and 2.11, which will be sufficient to demonstrate that the statistical file is valid.
3 Status Determinations	States often use statistical files for reporting status determinations when their system stores only the most recent status determination for each employer account and thus overwrites or overlays some status determinations. These files are often called "RQC files" because they were developed to provide a universe of determinations from which to derive the RQC sample. These States cannot perform an independent count from the database to validate the statistical file because the database will not contain records for all of the status determinations. Therefore, an independent count is not required for status determinations, because it is not possible to create such a count in States which use statistical files.
4 Accounts Receivable	All States must use a transaction history file or audit trail to correctly reconstruct payments (amounts liquidated), because only such files show the date that each payment was made. Transaction history files are also the source for receivable amounts established and amounts declared uncollectible in some States. There is only one source file for such transactions, so an independent count is not relevant. All States must use "employer quarter files" to reconstruct balances for reporting amounts removed and amounts outstanding at the end of the quarter. Some States use such balances for reporting amounts declared uncollectible. These balances are always captured as a "snapshot" at the end of the quarter from the database, so an independent count is not relevant.
5 Field Audits	States do not maintain more than one file with field audit results, thus an independent count is not relevant.

•Appendix D

INDEPENDENT COUNT SUMMARY SPREADSHEET

INTRODUCTION

When a State can only create the validation file from the same file it uses to produce the ETA 581 Report, then the reconstruction method will not produce meaningful results. In that case the State conducts an Independent Count as outlined in Appendix C. This appendix provides a spreadsheet to be used when entering the results of the Independent Count.

INDEPENDENT COUNT
SUMMARY SPREADSHEET

POPULATION	REPORT/ RECONSTRUC-TION DATA SOURCE	REPORTED COUNT	ADJUSTED VALIDATION COUNT	INDEPENDENT COUNT	% DIFFERENCES		ADJUSTED INDEPENDENT COUNT	PASS / FAIL	COMMENTS
					REPORTED COUNT/ INDEPENDENT COUNT	ADJUSTED VALIDATION COUNT/ INDEPENDENT COUNT			
1									
2									
3									
4									
5									

The independent count may legitimately be higher than the validation count if it includes observations in issue codes not specified on the RIV printout.
An adjusted independent count may be calculated in these circumstances and should be explained in the comments column.

Appendix E

VALIDATION WORKSHEETS

INTRODUCTION

Appendix E contains worksheets that States will use to record the results of data tests in Modules 1 and 2. There is a worksheet for each RV population. The first portion of the form is used to report the results of File Integrity Checks. The second portion summarizes the results of Data Element Validation, the review of sorted versions of the RV files. This latter information is also entered in the spreadsheets in Appendix B.

APPENDIX E

VALIDATION WORKSHEET 1
Active Employers

File Integrity Validation							
Step Type	**Sub-populations**	**Column #**	**Number of Cases Checked**			**Pass**	
FIV	1.1	All					
FIV	1.2	All					
Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	1.S.1	Employer Status	3A/1				
Sort	1.S.2	Employer Type	2A/1				
Sort	1.S.3	Employer Type	2B/1				
Sort	1.S.4	1. Liability Date (reopen) 2. Liability Date (initial)	3A/2				
Sort	1.A	EAN	2A/2				
Sort	1.B	EAN	2B/2				
Sort	1.C	Inactive/Terminated "as of" Date	3A/3				
Sort	1.D	1. Number of Liable Quarters 2. Sum of wages	3A/4				

NOTES: Validator should ensure, prior to beginning validation, that the appropriate reconstruction file and sort file pages are printed and assembled with this worksheet, RV spreadsheets, independent count spreadsheet (if applicable), and screens used in FIV. (See Module 2.)

APPENDIX E

VALIDATION WORKSHEET 2
Report Filing

File Integrity Validation				
Step Type	**Sub-populations**	**Column #**	**Number of Cases Checked**	**Pass**
FIV	2.1	All		
FIV	2.2	All		
FIV	2.3	All		
FIV	2.4	All		
FIV	2.5	All		
FIV	2.6	All		
FIV	2.7	All		
FIV	2.8	All		
FIV	2.9	All		
FIV	2.10	All		
FIV	2.11	All		
FIV	2.12	All		
FIV	2.13	All		
FIV	2.14	All		
FIV	2.15	All		
FIV	2.16	All		

APPENDIX E

VALIDATION WORKSHEET 2
Report Filing

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	2.S.1	Employer Type	2A/1				
Sort	2.S.2	Employer Type	2B/1				
Sort	2.S.3	Filing Status Indicator	8A/1				
Sort	2.S.4	Filing Status Indicator	8B/1				
Sort	2.S.5	Filing Status Indicator	8C/1				
Sort	2.S.6	Received Date	8A/2				
Sort	2.S.7	Received Date	8B/2				
Sort	2.S.8	Received Date	8C/2				
Sort	2.A	EAN	2A/2				
Sort	2.B	EAN	2B/2				
Sort	2.C	Final Assessment Date	8C/3				
Sort	2.D	Liability Date (Met Threshold)	8C/4				

APPENDIX E

VALIDATION WORKSHEET 2
Report Filing

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	2.E	Liability Date (Initial or Reopen)	8C/5				
Sort	2.F	Inactive/ Terminated "as of "Date	8C/6				
Sort	2.G	Suspended "as of" Quarter	8C/7				
Sort	2.H	Inactive/ Termination Processing Date	8C/8				

NOTES: Validator should ensure, prior to beginning validation, that the appropriate reconstruction file and sort file pages are printed and assembled with this worksheet, RV spreadsheets, independent count spreadsheet (if applicable), and screens used in FIV. (See Module 2.)

VALIDATION WORKSHEET 3
Status Determinations

File Integrity Validation							
Step Type	**Sub-populations**	**Column #**	**Number of Cases Checked**			**Pass**	
FIV	3.1	All					
FIV	3.2	All					
FIV	3.3	All					
FIV	3.4	All					
FIV	3.5	All					
FIV	3.6	All					
FIV	3.7	All					
FIV	3.8	All					
Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	3.S.1	Status Determination Type	11A/1				
Sort	3.S.2	Status Determination Type	11B/1				
Sort	3.S.3	Status Determination Type	11C/1				
Sort	3.S.4	Status Determination Type	11D/1				

VALIDATION WORKSHEET 3
Status Determinations *(continued)*

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	3.S.5	Time Lapse	12/1				
Sort	3.A	Status Determination Date	13/1				
Sort	3.B	1. Activation Processing Date 2. Reactivation Processing Date	11A/2				
Sort	3.C	Successorship Processing Date	11B/3				
Sort	3.D	Predecessor Account Number	11B/2				
Sort	3.E	Inactivation Processing Date	11C/2				
Sort	3.F	Termination Processing Date	11D/2				
TPS Acceptance Sample issues:							

NOTES: Validator should ensure, prior to beginning validation, that the appropriate reconstruction file and sort file pages are printed and assembled with this worksheet, RV spreadsheets, independent count spreadsheet (if applicable), and screens used in FIV. (See Module 2.)

VALIDATION WORKSHEET 4
Accounts Receivable

File Integrity Validation				
Step Type	**Sub-populations**	**Column #**	**Number of Cases Checked**	**Pass**
FIV	4.1	All		
FIV	4.2	All		
FIV	4.3	All		
FIV	4.4	All		
FIV	4.5	All		
FIV	4.6	All		
FIV	4.7	All		
FIV	4.8	All		
FIV	4.9	All		
FIV	4.10	All		
FIV	4.11	All		
FIV	4.12	All		
FIV	4.13	All		
FIV	4.14	All		
FIV	4.15	All		
FIV	4.16	All		

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	4.A	Transaction Date	23/1 24/1				

VALIDATION WORKSHEET 4
Accounts Receivable *(continued)*

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	4.B	Established Q/Date	22/1				
Sort	4.C	Transaction Type	22/2				
Sort	4.D	Amount Established in RQ	22/3				
Sort	4.E	Transaction Type	23/2				
Sort	4.F	Liquidated (Pay/Adj)	23/3				
Sort	4.G	Established Q/Date	24/2 25/1				
Sort	4.H	Employer RQ	24/3 25/4 26/3				
Sort	4.I	Employer RQ	24/4 26/2				
Sort	4.J	Due Date	24/5 25/6 26/5				
Sort	4.K	Due Date	24/6 26/4				
Sort	4.L	Transaction Type	24/7				
Sort	4.M	Uncollectible	24/8				

VALIDATION WORKSHEET 4
Accounts Receivable *(continued)*

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	4.N	Established Q/Date	25/2				
Sort	4.O	Employer RQ	25/3				
Sort	4.P	Due Date	25/5				
Sort	4.Q	Removed	25/7				
Sort	4.R	Established Q/Date	26/1				
Sort	4.S	Balance at end of RQ	26/6				

NOTES: Validator should ensure, prior to beginning validation, that the appropriate reconstruction file and sort file pages are printed and assembled with this worksheet, RV spreadsheets, independent count spreadsheet (if applicable), and screens used in FIV. (See Module 2.)

VALIDATION WORKSHEET 5
Field Audits

File Integrity Validation							
Step Type	**Sub-populations**	**Column #**	**Number of Cases Checked**			**Pass**	
FIV	5.1	All					
FIV	5.2	All					
FIV	5.3	All					
FIV	5.4	All					
Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	5.S.1	Employer Size	28A/1				
Sort	5.S.2	Employer Size	28B/1				
Sort	5.S.3	Change Audit	29A/1				
Sort	5.S.4	Change Audit	29B/1				
Sort	5.S.5	Audit Completion Date	30/1				
Sort	5.A	Total Wages Reconciliation	31E/1				

VALIDATION WORKSHEET 5
Field Audits *(continued)*

Data Element Validation							
Step Type	**Sort Number**	**Sort Key**	**Step/Rule in Handbook**	**# of Cases Checked**	**# of Errors**	**% of Errors**	**Pass or Fail**
Sort	5.B	Taxable Wages Reconcilia-tion	32E/1				
Sort	5.C	Contribu-tions Reconcilia-tion	33E/1				
TPS Acceptance Sample issues:							

NOTES: Validator should ensure, prior to beginning validation, that the appropriate reconstruction file and sort file pages are printed and assembled with this worksheet, RV spreadsheets, independent count spreadsheet (if applicable), and screens used in FIV. (See Module 2.)

WAGE ITEM VALIDATION WORKSHEET

Mode	Time Period	581 Count for Batch	Recount for Batch	Number of Errors	Percentage of Errors	Missing ID	Missing Amount	Double Count
Magnetic Tape								
Electronic Transfers								
Data Entry								
Scanning								

•Appendix F

ALTERNATIVE METHODOLOGY FOR DUPLICATE DETECTION

A. PURPOSE

States are encouraged to automate the validation process as much as possible. One method of accomplishing this is through the use of frequency distributions, which can be useful for report validation (Module 1.1), duplicate detection (Module 1.2). Frequency distributions are easily produced by general purpose data analysis software, such as SAS or SPSS. Frequency counts can also be produced using the Structured Query Language (SQL) programs that are components of relational database management systems (RDMS) such as Oracle, Ingres, Sybase, and Informix.

B. DUPLICATE DETECTION

Using frequency distributions for detecting duplicates provides several benefits over a sampling approach. All duplicates in the population will be detected in the frequency counts, whereas a sample selected from the population might not include all of the duplicate records. Also, the number of duplicates in all subpopulations can be identified, rather than estimating for the population as a whole.

The procedure consists of three tasks:

1. Count the number of records in the population that match each unique combination of data elements displayed in Figure 1.6. For this example, using population 3, the duplicate detection criteria are the Employer Account Number (EAN), Status Determination Date, Type of Status Determination and Predecessor Account Number.
2. Determine if the frequency count for any unique combination of data elements is greater than 1 (signaling a duplicate, triplicate, etc.).
3. Print the records for which duplicates have been detected and research to determine the cause for the duplicate and whether or not it is valid. (Population 3 may contain duplicate EANs which are valid since an employer may have more than one status determinations during a quarter. However, no individual determination should appear in the RV file more than once.)

Appendix G

SUMMARY OF DATA VALIDATION TASKS

The tables in this appendix offer a summary of the major tasks that states will need to accomplish as they complete data validation. The tables also indicate the types of staff that would typically have the prime responsibility for each task. Staff roles are divided among:

- ***Automated data processing (ADP) staff***, who have the primary responsibility for extracting data from the database to create the reconstruction files. They also sort and format those files so they are useful to validators.

- ***Validator***s, the end users who test the reconstructed data and then assess the validity of the information the State has reported on the ETA 581 report. Validators should work closely with ADP staff to determine the information that belongs in the reconstruction files.

- ***Managers*** are responsible for assuring that (1) the data validation process stays on track and (2) the data validation team has the resources it needs to meet the requirements of this handbook and the schedule set by the State.

This set of tables is intended to be a checklist that each State can use as a guide when developing its plan. With different priorities, reporting systems and staff capacities, each State will develop a different plan to suit its unique circumstances and preferences.

Finally, though we talk here about a single plan, many states will want to develop different plans for validating different populations. These plans may have different staff and schedules.

APPENDIX G

Task 1: Preparing for Data Validation	
Activity	**Roles**
Assemble data validation team.	Managers
Review handbook.	Validators, ADP
Attend training. Share training with staff who did not attend.	Validators, ADP
Review and update state-specific information in Module 3 of the handbook. Send US DOL and its contractors a copy of the module with any needed changes clearly marked.	ADP, with help from validators
Develop a data validation plan with: 1. Schedule for completing data processing and validation review ***for each population.*** ***2.*** Staff assignments for each step in the data validation process.	Managers, validators, ADP

Task 2: Creating Reconstruction (RV) Files	
Activity	**Roles**
Review specifications, product requirements, and schedules.	ADP, Validators
Convert handbook specifications into programming specifications.	ADP
Develop "capture" programs if needed.	ADP
Develop file extract programs.	ADP
Develop programs to format extracted files for reviewers.	ADP
Review test output for consistency with handbook requirements.	Validators
Modify programs to correct any problems identified.	ADP
Schedule data validation extract programs to run at the same time ETA 581 programs are executed. Also arrange to print any screens needed for Module 2 at the same time.	Managers, ADP, Validators

APPENDIX G

Task 3: Duplicate Detection	
Activity	**Roles**
Following the specifications in Table 1.2.1, Duplicate Detection Criteria, determine the sample type and size required for each population	ADP, Validators
Select a sample of transactions (or for population 3 a sample of duplicate EANs).[a]	ADP
Print out the records for employers in the sample after sorting them by EAN.	ADP
Review the EANs in the sample for population 3 to determine if they are true duplicates. Use the duplicate detection criteria in Table 1.2.1 to guide the review.	Validators
Enter the duplicate detection findings in the RV spreadsheet.	Validators

[a] Employer Account Number

Task 4: File Integrity Validation	
Activity	**Roles**
Secure a printout of the first page of the RV printout for each subpopulation. Mark the first two records.	Validators, ADP
In advance, have necessary query screens produced at same time reconstruction file is created.	Validators or ADP
Following the steps indicated in the work sheets in Appendix A and described in Module 3, review and validate every item (column) on the printout for the two selected records.	Validators
Record the results on the top half of the worksheet**.**	Validators
If invalid data were used in the creation of the reconstruction file, correct the file and begin this task over again. This is obviously a step that should be done well before the scheduled date for the actual validation.	ADP
Conduct Item Validation for corrected versions of the RV file.	Validators

Task 5: Data Element Validation Sorting the Reconstruction Files	
Activity	**Roles**
Identify Sort Requirements from Appendix A, Part II	ADP, Validators
Sort RV files as indicated in the Appendix. This will produce several new files for each subpopulation. Maintain same headings and record layout as in original file. Change only the order of transactions and the title of the printout.	ADP
Review first test of sorted files. Make any revisions needed.	ADP, Validators
Review first and last page of each sort file. The review should be guided by the steps listed in the sort criteria, and discussed in the Data Element Validation instructions within Module 3. Print out and retain the first and last pages of each sorted version of the file.	Validators
Review additional transactions for sorts 1.C and 3.S.5	Validators
Calculate percentage error and enter results on Worksheets from Appendix E.	Validators

Task 6: Recording the Results	
Activity	**Roles**
Enter report counts and validation counts in the spreadsheets found in Appendix B. This will normally be done manually, although states have the option to develop processes to automatically transfer information into spreadsheets.	Validators/managers or ADP
The spreadsheet will calculate a Pass/Fail verdict on the accuracy of each validated item.	Validators/managers or ADP
Submit the findings as documented on the spreadsheets.	Managers